frontiers | Research Topics

Engineering microalgal chassis cells

Edited by
Xuefeng Lu, Martin Hagemann, Jin Liu, Pratyoosh Shukla and Xiaoming Tan

Published in
Frontiers in Microbiology

ISSN 1664-8714
ISBN 978-2-8325-2963-8
DOI 10.3389/978-2-8325-2963-8

About Frontiers

Frontiers is more than just an open access publisher of scholarly articles: it is a pioneering approach to the world of academia, radically improving the way scholarly research is managed. The grand vision of Frontiers is a world where all people have an equal opportunity to seek, share and generate knowledge. Frontiers provides immediate and permanent online open access to all its publications, but this alone is not enough to realize our grand goals.

Frontiers journal series

The Frontiers journal series is a multi-tier and interdisciplinary set of open-access, online journals, promising a paradigm shift from the current review, selection and dissemination processes in academic publishing. All Frontiers journals are driven by researchers for researchers; therefore, they constitute a service to the scholarly community. At the same time, the *Frontiers journal series* operates on a revolutionary invention, the tiered publishing system, initially addressing specific communities of scholars, and gradually climbing up to broader public understanding, thus serving the interests of the lay society, too.

Dedication to quality

Each Frontiers article is a landmark of the highest quality, thanks to genuinely collaborative interactions between authors and review editors, who include some of the world's best academicians. Research must be certified by peers before entering a stream of knowledge that may eventually reach the public - and shape society; therefore, Frontiers only applies the most rigorous and unbiased reviews. Frontiers revolutionizes research publishing by freely delivering the most outstanding research, evaluated with no bias from both the academic and social point of view. By applying the most advanced information technologies, Frontiers is catapulting scholarly publishing into a new generation.

What are Frontiers Research Topics?

Frontiers Research Topics are very popular trademarks of the *Frontiers journals series*: they are collections of at least ten articles, all centered on a particular subject. With their unique mix of varied contributions from Original Research to Review Articles, Frontiers Research Topics unify the most influential researchers, the latest key findings and historical advances in a hot research area.

Find out more on how to host your own Frontiers Research Topic or contribute to one as an author by contacting the Frontiers editorial office: frontiersin.org/about/contact

Engineering microalgal chassis cells

Topic editors

Xuefeng Lu — Qingdao Institute of Bioenergy and Bioprocess Technology, Chinese Academy of Sciences (CAS), China

Martin Hagemann — University of Rostock, Germany

Jin Liu — Peking University, China

Pratyoosh Shukla — Banaras Hindu University, India

Xiaoming Tan — Hubei University, China

Citation

Lu, X., Hagemann, M., Liu, J., Shukla, P., Tan, X., eds. (2023). *Engineering microalgal chassis cells*. Lausanne: Frontiers Media SA. doi: 10.3389/978-2-8325-2963-8

Table of contents

frontiers | Frontiers in Microbiology

TYPE Editorial
PUBLISHED 26 June 2023
DOI 10.3389/fmicb.2023.1237999

OPEN ACCESS

EDITED AND REVIEWED BY
William James Hickey,
University of Wisconsin-Madison, United States

*CORRESPONDENCE
Xuefeng Lu
✉ lvxf@qibebt.ac.cn

RECEIVED 10 June 2023
ACCEPTED 16 June 2023
PUBLISHED 26 June 2023

CITATION
Lu X, Hagemann M, Liu J, Shukla P and Tan X (2023) Editorial: Engineering microalgal chassis cells. *Front. Microbiol.* 14:1237999. doi: 10.3389/fmicb.2023.1237999

Editorial: Engineering microalgal chassis cells

Xuefeng Lu[1]*, Martin Hagemann[2], Jin Liu[3], Pratyoosh Shukla[4] and Xiaoming Tan[5]

[1]Qingdao Institute of Bioenergy and Bioprocess Technology, Chinese Academy of Sciences (CAS), Qingdao, China, [2]Institute of Biosciences, Department of Plant Physiology, University of Rostock, Rostock, Germany, [3]Laboratory for Algae Biotechnology & Innovation, College of Engineering, Peking University, Beijing, China, [4]Enzyme Technology and Protein Bioinformatics Laboratory, School of Biotechnology, Institute of Science, Banaras Hindu University, Varanasi, India, [5]State Key Laboratory of Biocatalysis and Enzyme Engineering, School of Life Sciences, Hubei University, Wuhan, China

KEYWORDS

microalgae, chassis cell, synthetic biology, metabolic engineering, photosynthetic production

Editorial on the Research Topic
Engineering microalgal chassis cells

Introduction

Microalgae play an important role in the oxidation and carbon-reduction process of the Earth's atmosphere, and they currently remain as the essential source of primary productivity to maintain the stability and development of the biosphere. In the context of increasing global energy and environmental crises, microalgae are also valued as highly promising microbial photosynthetic platforms. Using microalgae for the one-step conversion of CO_2 and solar energy to biofuels, bio-based chemicals and biomedical products is an attractive new technological paradigm for achieving carbon-neutral sustainable development. Traditionally, the development of microalgal biotechnology relied on the mining and utilization of natural algal species resources, while the rapid development of synthetic biology and metabolic engineering technologies is gradually changing the paradigm of this field. By modifying the natural pathways and introducing heterologous modules, the microalgal metabolism network can be remodeled to relocate the organic carbon flow fixed through Calvin cycle to artificial metabolic pathways for the synthesis of desired products. Meanwhile, by combination of the tailored metabolic and physiological functions of microalgal cells with material- and electrochemical-devices, the application scenario of microalgal biotechnology have been further expanded to the areas of biomedicine and biotherapy, biophotovoltaics and biofuel cells, and aerospace.

However, from the perspective of the overall socio-economic and industrial development level, the achievements mentioned above are generally still the "art products" from "studios" or "small workshops" guided by empirical design. In the R & D processes, a majority of microalgal research and engineering manipulations can hardly be performed in the high-throughout, automated bio-foundry platforms. Regarding the outputs, the efficacy and robustness of microalgal cell factories and biocatalytic systems are usually difficult to compete with the classical heterotrophic systems, and even not yet economically feasible. To accelerate the development and industrialization of advanced microalgal biotechnologies, which satisfies the requirements of both sustainability and economic-feasibility, breakthroughs in approaches and platforms are required in the existing

microalgal research and engineering system, and the development of universal chassis cells will play a key role. Toward the development and application of the new generation microalgal chassis cells, our understandings on the function and particularly regulation of microalgal genetics, physiology, and metabolism are yet to be expanded, the toolboxes for microalgal genome engineering need to be updated and enriched, and efficiency and robustness of current microalgal cell factories should be significantly improved. The publications in this Research Topic focus on addressing issues for *Engineering microalgal chassis cells*.

Currently, microalgae used for synthetic biology and metabolic engineering research are mostly unicellular species, mainly because of the morphology characteristics facilitating convenient genetic modifications and physiological and biochemical assays. However, unicellular microalgal strains would face difficulties in harvesting, poor resistance to adversity, and predation by protozoa, when being cultivated in outdoor and industrial conditions. In recent years, filamentous microalgae have gained attention due to their ideal industrial properties, and the development of filamentous microalgae chassis cells is important for enriching the microalgae synthetic biotechnology system. Bozan et al. sequenced a filamentous diazotrophic cyanobacterium *Tolypothrix* sp. PCC 7712, which is suited for cultivation with the capillary biofilm reactor (CBR) to reach high biomass production rates. Surprisingly, although PCC 7712 was identified to possess top performance in CBRs cultivation, its genome sequence has a high similarity to that of another species *Tolypothrix* sp. PCC 7601 (*Fremyella diplosiphon*), and the detailed comparative genomic analysis revealed that the physiological differences between the two strains might be resulted from the deviations on gene composition and arrangement. Understandings on genome sequence and characteristics of PCC 7712 would also facilitate genomic modification and metabolic engineering of this strain in future, leading to robust photosynthetic biomanufacturing processes suitable for industrial applications.

Compared to the classical heterotrophic microorganisms, tools for understanding, engineering and harnessing microalgae, are still limited. Developing a suited biotechnology toolbox would be a prerequisite for engineering and applying microalgal chassis cells. Transposon insertion mutagenesis is an important approach used to explore gene functions in microalgae, and Hu et al. reviewed the application of transposon insertion site sequencing method aiming to provide theoretical and technical support when using this strategy. The availability of appropriate plasmid vectors sets is of great significance for introducing and expressing heterologous proteins in microalgae. Sakamaki et al. reported the characterization of a cyanobacterial Rep protein (a replication initiating factor) in the model cyanobacterium *Synechocystis* sp. PCC 6803, which exhibits high autonomous replication activity in multiple cyanobacteria, and established a robust expression vector with this protein. Kaltenbrunner et al. discovered that the copy number of pSYSA defense plasmid in *Synechocystis* sp. PCC 6803 is positively related with the expression level of endoribonuclease E, and identified another protein determining the stability of this plasmid. The information and devices given in the above research would benefit the development of novel shuttle vectors for genetic engineering of cyanobacteria. Besides the "hardware" facilitating direct genetic manipulations, genome metabolic model as a "software" is playing an increasingly important role in microbial synthetic biology and metabolic engineering research. Inwongwan et al. summarized the progress and challenge on developing *Euglena gracilis* metabolic network model, and prospected the future application of this approach to guide precise metabolic engineering. Santos-Merino, Gargantilla-Becerra et al. reported the development of an updated genome-scale model of a freshwater cyanobacterium *Synechococcus elongatus* PCC 7942, and evaluated the potentials of this chassis strain for production of α-linolenic acid. Such a model could also be expected to benefit the development of photosynthetic cell factories of many other metabolites.

Microalgal cells would be challenged by multiple environmental stress factors, and thus it is important to improve the cellular robustness and fitness of the microalgal chassis cells, aiming to achieve rapid and stable carbon fixation, cell growth, and photosynthetic production in diverse conditions. Cantrell et al. reported that disruption of glycogen cycle or sucrose cycle in PCC 6803 led to redirected cellular energy for faster growth under high light conditions, which would also relieve the cell growth retardation caused by the high light stress in outdoor conditions. Dong et al. introduced a heterologous pathway for ectoine synthesis in PCC 7942 and improved salt tolerance of this strain, which might improve the performance of PCC 7942 cell factories when using sea-water for cultivation. Besides harmful environmental factors, the accumulation of products or intermediate metabolites would also cause inhibition on cell growth and metabolism of microalgal cells. Zhang H. et al. reported that expression of tardigrade disordered proteins (TDPs) from water bears could regulate the tolerance of PCC 6803 toward biofuels and metal ions. Through modification of the native metabolism network or introduction of heterologous modules, the physiological adaptation space of microalgal chassis cells and cell factories were effectively expanded, and more robust photosynthetic biosynthesis process could be expected.

The remaining publications in this topic focused on engineering microalgae for photosynthetic production of diverse natural or non-natural metabolites. Zhu et al. summarized the trends and progress on developing microalgal cell factories for lipids production and accumulation through metabolic and process optimizations. In another review article, Santo-Merino, Yun et al. focused on photosynthetic production of sucrose with cyanobacteria; the authors gave a comprehensive review of the current understandings about sucrose metabolism network and regulatory mechanism, summarized the genetic manipulation efforts to optimize sucrose titers, and prospected the trends in developing artificial consortia based on cyanobacterial sucrose secretion. Aiming to guide the optimization the carotenoids production microalgae, Zhang Z. et al. deciphered the role of cryptochromes as photoreceptors to mediate the blue light induced biosynthesis of carotenoids in *Phaeodactylum tricornutum*, and Jiang et al. reported that the overexpression of a plastid lipid-associated protein (PAP) led to optimization of xanthophyll synthesis and accumulation in the same strain. In another study, Opel et al. reported the efforts to develop a synthetic hydrogen sensor in cyanobacteria by introducing the *Cupriavidus nicators* sourced oxygen-tolerant regulatory hydrogenase, which showed distinct H_2 oxidation activity in PCC 6803. Although the complete

H_2 sensing cascade is yet to be established in cyanobacteria, a functional circuit was successfully constructed in *Escherichia coli* utilizing the associated two-component system. When the functional H_2 biosensor system was finally constructed and implemented in cyanobacteria in future, it would undoubtedly be an important contribution to the development and optimization of hydrogen-producing cell factories.

Overall, the publications in this topic presented the efforts, progresses, and challenges to engineer a new generation of microalgae chassis cells, which would lay a foundation to develop more efficient and robust microalgal biotechnologies and pave the way for more sustainable manufacturing routes of energy and materials.

Author contributions

XL drafted this editorial and all authors are participated in the revision. All authors approved the final version.

Funding

The funding support from the National Key R&D Program of China (2021YFA0909700), the National Natural Science Foundation of China (32271484), and Shandong Taishan Scholarship (to XL) are appreciated. PS acknowledges Science and Engineering Research Board (SERB), GoI (grant no. CRG/2021/001206), and the Faculty Incentive Grant by Institute of Eminence (IoE) Scheme by BHU, Varanasi (no. R/Dev/D/IoE/Seed & Incentive/2022-23/50024).

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

frontiers | Frontiers in Microbiology

TYPE Mini Review
PUBLISHED 10 October 2022
DOI 10.3389/fmicb.2022.1024441

OPEN ACCESS

EDITED BY
Pratyoosh Shukla,
Banaras Hindu University,
India

REVIEWED BY
Dillirani Nagarajan,
National Cheng Kung University, Taiwan
Xiao-Man Sun,
Nanjing Normal University, China
Pau Loke Show,
University of Nottingham Malaysia Campus, Malaysia

*CORRESPONDENCE
Xufeng Liu
xufeng.liu@kemi.uu.se
Peter Lindblad
peter.lindblad@kemi.uu.se

SPECIALTY SECTION
This article was submitted to Microbiotechnology, a section of the journal Frontiers in Microbiology

RECEIVED 21 August 2022
ACCEPTED 26 September 2022
PUBLISHED 10 October 2022

CITATION
Zhu Z, Sun J, Fa Y, Liu X and Lindblad P (2022) Enhancing microalgal lipid accumulation for biofuel production. *Front. Microbiol.* 13:1024441. doi: 10.3389/fmicb.2022.1024441

Enhancing microalgal lipid accumulation for biofuel production

Zhi Zhu[1,2,3,4], Jing Sun[1], Yun Fa[3], Xufeng Liu[4]* and Peter Lindblad[4]*

[1]The Key Laboratory of Biotechnology for Medicinal Plants of Jiangsu Province, School of Life Sciences, Jiangsu Normal University, Xuzhou, China, [2]State Key Laboratory of Biocatalysis and Enzyme Engineering, School of Life Sciences, Hubei University, Wuhan, China, [3]CAS Key Laboratory of Bio-Based Materials, Qingdao Institute of Bioenergy and Bioprocess Technology, Chinese Academy of Sciences, Qingdao, China, [4]Microbial Chemistry, Department of Chemistry-Ångström Laboratory, Uppsala University, Uppsala, Sweden

Microalgae have high lipid accumulation capacity, high growth rate and high photosynthetic efficiency which are considered as one of the most promising alternative sustainable feedstocks for producing lipid-based biofuels. However, commercialization feasibility of microalgal biofuel production is still conditioned to the high production cost. Enhancement of lipid accumulation in microalgae play a significant role in boosting the economics of biofuel production based on microalgal lipid. The major challenge of enhancing microalgal lipid accumulation lies in overcoming the trade-off between microalgal cell growth and lipid accumulation. Substantial approaches including genetic modifications of microalgal strains by metabolic engineering and process regulations of microalgae cultivation by integrating multiple optimization strategies widely applied in industrial microbiology have been investigated. In the present review, we critically discuss recent trends in the application of multiple molecular strategies to construct high performance microalgal strains by metabolic engineering and synergistic strategies of process optimization and stress operation to enhance microalgal lipid accumulation for biofuel production. Additionally, this review aims to emphasize the opportunities and challenges regarding scaled application of the strategic integration and its viability to make microalgal biofuel production a commercial reality in the near future.

KEYWORDS

microalgae, lipid enhancement, biofuel production, metabolic engineering, process optimization

Introduction

Biofuel is a form of energy which captures solar energy as chemical energy in the bonds of biologically produced materials (Srivastava et al., 2020). As one of the most important study aspects in exploitation and application of the renewable energy, biofuel plays a significant role in dealing with the increasing demand of energy and the deteriorating

environmental pollution problems (Medipally et al., 2015; Ong et al., 2020; Peng et al., 2020). Compared with others, lipid-based biofuels have been attracting extensive attention due to the higher energy density, better infrastructure compatibility and greater application flexibility (Wang et al., 2022).

Unicellular microalgae are photoautotrophic organisms which grow like photosynthetic plants while lacking the complex cell structures of higher plants (Slade and Bauen, 2013). Microalgae have been considered as one of the most promising alternative sustainable feedstocks for producing lipid-based biofuels due to their higher lipid accumulation capacity, higher growth rate and higher photosynthetic efficiency compared to the traditional plants (Chisti, 2007; Chu, 2017; Anto et al., 2020; Wang et al., 2020). In addition, microalgae provide proteins that can be used as feed source for animals (Amorim et al., 2021). Some of them can also produce high value biologically active compounds like some antioxidant pigments (Markou and Nerantzis, 2013). Microalgae have been displayed greater sustainable and commercial advantages as feedstock for biofuels production (Harun et al., 2010; Saranya and Shanthakumar, 2021).

Microalgae could offer great prospect for biofuel exploitation. However, the process is still not carbon neutral and commercially viable because of the high production cost (Behera et al., 2021; Brar et al., 2021). Enhancing microalgal lipid accumulation could improve the economic feasibility of the biofuel production. Several recent reviews have summarized genetic and metabolic engineering approaches and/or cultivation regulating strategies for enhancing microalgal lipid accumulation or productivity, but a very few discussed these strategies all together for achieving high lipid production with more focus on the trade-off between microalgal cell growth and lipid accumulation (Chu, 2017; Sun et al., 2019; Khan and Fu, 2020; Shokravi et al., 2020; Brar et al., 2021).

The focus of this review is thus to highlight the advancements and emerging approaches towards achieving enhancement of microalgal lipid accumulation for biofuel production on the basis of the trade-off between microalgal cell growth and lipid accumulation. The scope of present work covers genetic manipulations of microalgal strains and optimizations of microalgal cultivation systems, along with their challenges.

Construction of high-efficient lipid producing microalgae

There are native biological routes for biofuel molecules in some of the natural strains of microalgae. Figure 1 gives a schematic overview of lipid metabolic pathways in microalgae. Microalgal lipid metabolic pathways are mainly based on fatty acid *de novo* biosynthesis pathway and triacylglycerol (TAG) synthesis route. Fatty acid *de novo* biosynthesis pathway in microalgae occurs in chloroplast. As primary substrate of fatty acid *de novo* biosynthesis, acetyl-CoA is carboxylated to malonyl-CoA generating saturated fatty acids, which go through further desaturation and elongation forming unsaturated fatty acids catalyzed by complex fatty acid synthases (FAS; Moffett et al., 2020; Behera et al., 2021). TAG synthesis pathway in endoplasmic reticulum has been proposed to be composed of three sequential transfers of acyl group from acyl-CoA to glycerol-3-phosphat (Brar et al., 2021; Mulgund, 2022). The natural production was generally low, restricting the industrialized production and commercialized development of the microalgal biofuels. Metabolic engineering is one of the most important research fields of biotechnology innovation, which helps to modify metabolic pathways to trigger the productions of the target biofuel metabolites.

Metabolic engineering of fatty acid *de novo* biosynthesis pathway

Malonyl-CoA: acyl carrier protein transacylase (MCAT) is responsible for the conversion of malonyl-CoA to malonyl acyl carrier protein, which is the first committed step of fatty acid biosynthesis, followed by the reduction-dehydration-reduction reaction cycle. Overexpression of MCAT was shown to increase the accumulation of the fatty acid (Lei et al., 2012; Tian et al., 2013). For example, the MCAT overexpressed *Schizochytrium* obtained 110.5 g/l total lipid in the fed-batch cultivation, which was 39.6% higher than that in cells of the wild strain. In addition, the production of polyunsaturated fatty acid was also increased when overexpressing MCAT (Li et al., 2018). Overexpression of MCAT in oleaginous microalga *Nannochloropsis oceanica* elevated the lipid content to 42.9% of the dry cell weight, leading to 36.0% higher content than that in cells of the wild strain (Chen et al., 2017).

Acetyl-CoA carboxylase (ACCase) is responsible for the conversion of acetyl-CoA to malonyl-CoA, making malonyl-CoA to enter the fatty acid biosynthesis pathway (Figure 1). Many studies have shown that upregulation of ACCase improve the biosynthesis of the fatty acid (Sun et al., 2019). Increasing ACCase expression by genetic engineering to overproduce fatty-acid based biofuels has been widely employed in model microorganisms such as *Escherichia coli* and *Saccharomyces cerevisiae* (Das et al., 2020). Gomma et al. were the first to report overexpressed ACCase for improving fatty acid biosynthesis in microalgae. The total fatty acid production of ACCase overexpressed mutant was increased by 60% compared to in cells of the wild type (Gomma et al., 2015).

Glucose-6-phosphate dehydrogenase (G6PD) is involved in the biosynthesis of NADPH in pentose phosphate pathway, which plays an important role in maintaining the reducing power and redox homeostasis. Xue et al. constructed a G6PD overexpressed mutant of *Phaeodactylum tricornutum* in which both the transcript abundance and enzyme activity of G6PD were increased as a result of enhancement of NADPH (Xue et al., 2017). The lipid content reached 55.7% of dry cell weight, 2.7-fold higher than that in cells of the wild type (Xue et al., 2017). By enhancing the reducing power supply, overexpression of G6PD in microalgae can

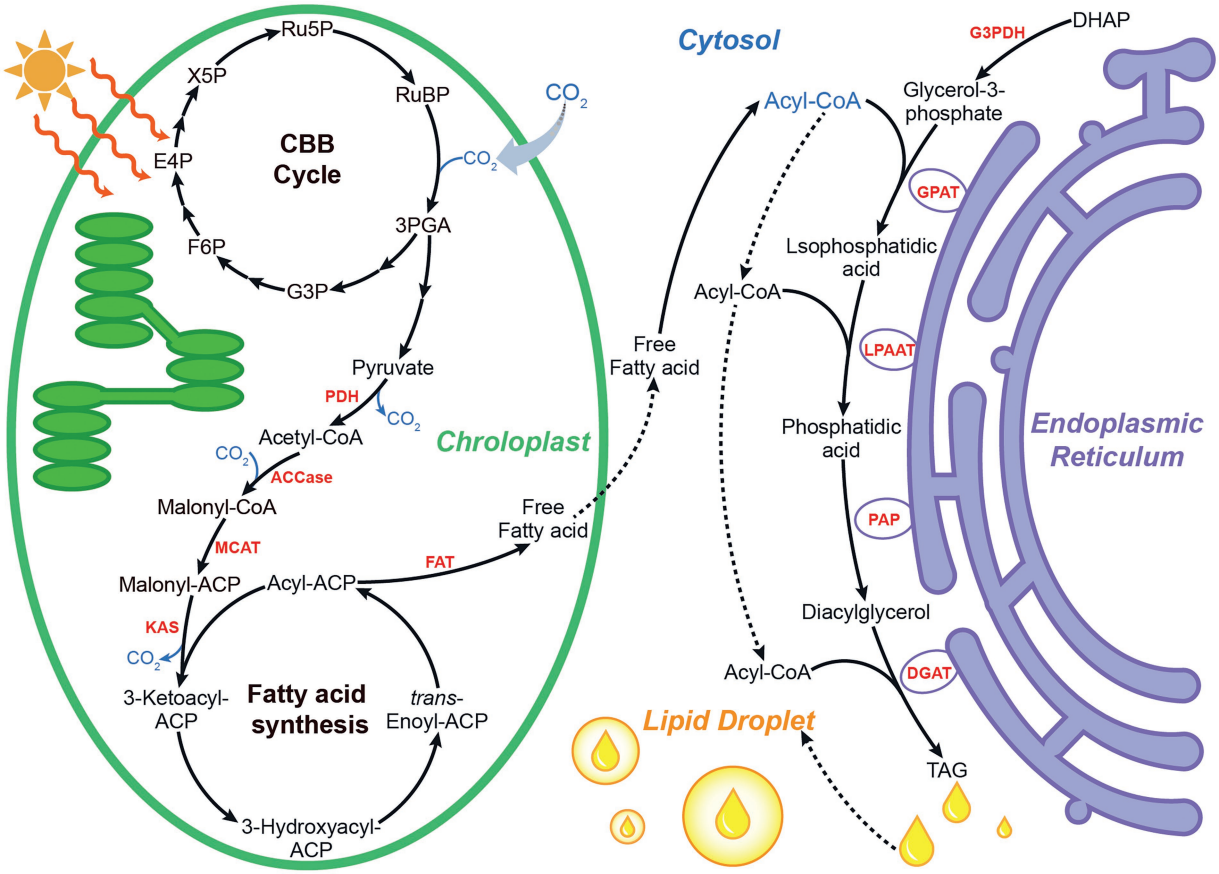

FIGURE 1
Schematic overview of metabolic pathways involved in lipid synthesis in microalgae. 3PGA, 3-phosphoglycerate; ACCase, acetyl-CoA carboxylase; DGAT, diacylglycerol acyltransferase; DHAP, dihydroxyacetone phosphate; E4P, erythrose-4-phosphate; F6P, fructose-6-phosphate; FAT, fatty acyl-ACP thioesterase; G3P, glyceraldehyde-3-phosphate; G3PDH, gycerol-3-phosphate dehydrogenase; GPAT, glycerol-3-phosphate acyltransferase; KAS, 3-ketoacyl-ACP synthase; LPAAT, lysophosphatidic acid acyltransferase; MCAT, malonyl-CoA: acyl carrier protein transacylase; PAP, phosphatidic acid phosphatase; PDH, pyruvate dehydrogenase complex; Ru5P, ribulose-5-phosphate; RuBP, ribulose-1,5-bisphosphate; TAG, triacylglycerols; X5P, xylulose-5-phosphate.

significantly improve the lipid accumulation, illustrating that G6PD may be a promising metabolic engineering target for efficient microalgal lipid production. Overexpression of malic enzyme was also proved playing a significant role in enhancing neutral lipid production of *P. tricornutum* through the additional supply of NADPH (Zhu et al., 2018).

Metabolic engineering of triacylglycerol synthesis pathway

In microalgae, glycerol-3-phosphate dehydrogenase (G3PDH) catalyzes the conversion of dihydroxyacetone phosphate (DHAP) to glycerol-3-phosphate in the cytosol, followed by glycerol-3-phosphate acyltransferase (GPAT) which catalyzes the conversion of glycerol-3-phosphate to lysophosphatidic acid, lysophosphatidic acid acyltransferase (LPAAT) the conversion of lysophosphatidic acid to phosphatidic acid, phosphatidic acid phosphatase (PAP) the conversion of phosphatidic acid to diacylglycerol, and finally diacylglycerol acyltransferase (DGAT) the conversion of diacylglycerol to triacylglycerol (TAG) in the endoplasmic reticulum (Korkhovoy and Blume, 2013; Figure 1). Hsieh et al. adopted a multiple gene expression strategy to elevate the lipid accumulation of microalgae (Hsieh et al., 2012). The coordinated overexpression of G3PDH, GPAT, LPAAT, PAP, and DGAT from *Saccharomyces cerevisiae* and/or *Yarrowia lipolytica* increased the lipid production of *Chlorella minutissima* 2-fold compared to in cells of the wild type (Hsieh et al., 2012). Wang et al. multi-overexpressed homologous GPAT and LPAAT in *Phaeodactylum tricornutum*, leading to 2.3-fold higher TAG content (with nitrogen stress condition) than that in cells of the wild strain (Wang et al., 2018). Compared with other TAG biosynthesis-related genes, the genes encoding GPAT and DGAT may be more effective targets for harnessing lipid accumulation in the TAG biosynthesis. Zou et al. also provided a multiple gene expression platform for manipulation of complex metabolic nodes. Coordinated expression of homologous GPAT and DGAT increased the lipid content of *Phaeodactylum tricornutum* by

2.6-fold than that of wild type (Zou et al., 2018). Zulu et al. also heterologously co-expressed DGAT from yeast and oleosin (lipid droplet stabilizing protein) from plant in *Phaeodactylum tricornutum* resulting in a 3.6-fold increased TAG content compared to in cells of the wild strain (Zulu et al., 2017). With respect to conventional single gene construction, Niu et al. reported that overexpression of homologous GPAT alone made *P. tricornutum* to produce twise as much neutral lipids compared to in wild type cells (Niu et al., 2016). Chen et al. constructed a genetically engineered strain of *Scenedesmus obliquus* harboring a DGAT gene from *Chlamydomonas reinhardtii*, which was successfully cultured in a 40l tubular photobioreactor. The lipid content of this recombinant strain reached 12.3% of dry cell weight, 128% higher than in cell of the wild strain (Chen et al., 2016). Collectively, metabolic engineering of key enzymes in the TAG biosynthesis pathway may be a promising strategy for microalgal lipid accumulation as required.

In addition to metabolic engineering for increasing the quantity of microalgal lipids, it is also rational to improve the quality of the producing lipids. Some efforts have been made on improving the lipids quality in terms of manipulating the degree of fatty acid unsaturation and the length of fatty acid carbon chain (Norashikin et al., 2018; Haslam et al., 2020; Wang et al., 2021).

Metabolic engineering of competitor pathways

Knockout or knockdown of key enzymes of competing pathways to allow more carbon flux to be channeled toward the target products is widely considered as an effective strategy to improve the production of the target products. Microalgae accumulate starch (carbohydrate) and together with lipids they are the two primary carbon storage metabolites under stressful conditions (Takeshita et al., 2014; Li T. T. et al., 2015). The metabolism of starch and lipids are highly related, and glyceraldehyde-3-phosphate (G3P) is their common precursor (Ran et al., 2019). Blocking this competitive starch photosynthetic pathway channeling the carbon flux toward lipid biosynthesis may represent an effective strategy to overproduce the target lipid. For instance, a *Chlamydomonas* starchless mutant was found to accumulate 10-fold more cellular TAG compared to that in cells of wild type when the cultures were transferred to a high light intensity and nitrogen-less medium (Li et al., 2010). de Jaeger et al. reported that the total fatty acid productivity of a *Scenedesmus obliquus* starchless mutant increased 41% compared to in wild type cells and the TAG yield reached 49.4% of the dry cell weight under nitrogen deficiency stress condition (de Jaeger et al., 2014).

In microalgae, tricarboxylic acid (TCA) cycle and *de novo* fatty acid biosynthesis pathway are also two competitive pathways of carbon storage, sharing the common mid-metabolite phosphoenolpyruvate (PEP). Phosphoenolpyruvate carboxylase (PEPC) is responsible for the conversion of PEP to oxaloacetate (OAA), making OAA to enter into TCA cycle. Disruption of TCA cycle may be an effective approach for carbon partitioning toward lipid accumulation. A PEPC down- regulated strain of *Chlamydomonas reinhardtii* was developed in which the maximal lipid content and productivity increased by 74.4 and 94.2%, respectively, compared in wild type cells (Kao and Ng, 2017).

Inhibition of competitive lipid catabolism pathway is another practical metabolic engineering strategy targeting overall lipid accumulation. For example, Trentacoste et al. developed transgenic strains of *Thalassiosira pseudonana* through targeted knocking down of multifunctional lipase/phospholipase/lysophosphatidic acyltransferase, resulting in 3.3-fold higher lipid yield than that of wild type cells at exponential phase and 4.1-fold higher lipid yield than under silicon deficiency stress condition without affecting cell growth (Trentacoste et al., 2013).

Optimization of nutrient conditions

It is important to obtain high-efficient microalgae for industrialized biofuel production. To realize the full potential of the high-efficient microalgae, it is essential to optimize the process of their growth and the products production (Figure 2). Generally, the guideline of optimizing nutrient conditions is to investigate nutrients uptake from the medium and the metabolic routes of the nutrients, then to confirm the effect of different nutrient conditions on microalgal cell growth and metabolites distribution, before finally to further optimize the nutrient conditions for improving the production of biomass and/or target products (Du et al., 2006; Ramirez-Lopez et al., 2016). Targeting the overall biofuel production based on microalgae, the major challenge of the optimization of culture conditions, as well as the genetic modification of the microalgal host strain, lies in the trade-off between microalgal cell growth and lipid accumulation (Singh et al., 2016). The content and the supplying mode of the key ingredient of nutrients, such as carbon source or nitrogen source, play significant roles in the cell growth and the lipid accumulation are summarized in Figure 2.

Carbon source

In photoautotrophic cultivation mode, microalgae utilize CO_2 and $NaHCO_3$ as inorganic carbon source to produce organic metabolites through photosynthesis. While in heterotrophic or mixotrophic mode, some microalgae assimilate organic molecules such as glucose or glycerol as carbon source and /or energy (Markou et al., 2014; Saranya and Shanthakumar, 2021). Tang et al. reported that the photosynthetic lipid content of both of *Scenedesmus obliquus* and *Chlorella pyrenoidosa* increased with increasing levels of environmental CO_2 (Tang et al., 2011). The maximum lipid content of *S. obliquus* reached 24.4% of dry cell weight under 50% CO_2 condition, 61% higher than that of 0.03%

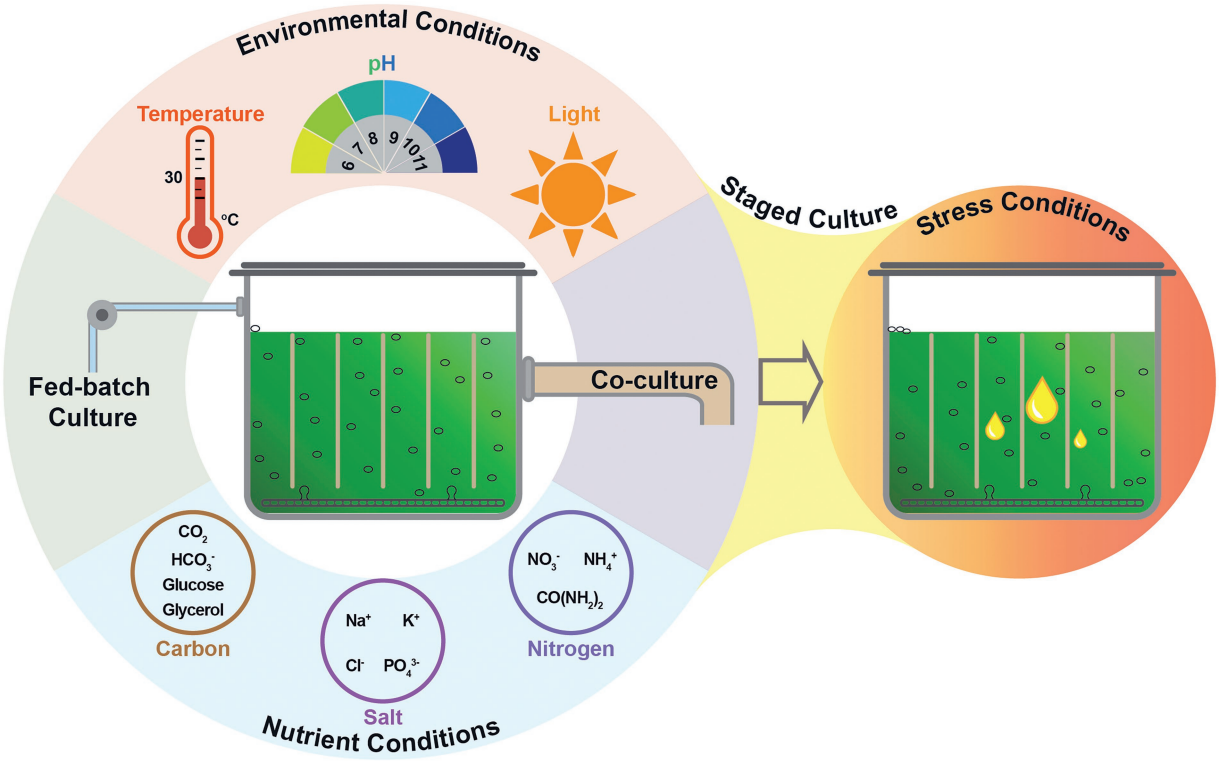

FIGURE 2
Systematic strategies of cultivation process tailoring microalgae for enhanced lipid production.

CO_2 condition (simulating ambient atmosphere). The maximum lipid content of *C. pyrenoidosa* reached 26.8% of dry cell weight under 50% CO_2 condition, 28% higher than that of 0.03% CO_2 condition (Tang et al., 2011). Compared with photoautotrophic cultivation, the maximum lipid content of *Chlorella sorokiniana* increased 2.4-fold with 20 g/l glucose under heterotrophic cultivation and 3.9-fold with 8 g/l glucose under mixotrophic cultivation conditions (Li et al., 2014). Laraib et al. obtained 137.43 ± 13.3 mg/l/d of biomass productivity and 39% of lipid content during the mixotrophic cultivation of *Chlorella vulgaris* when molasses utilized as additional carbon source, respectively 1.5-fold and 2-fold higher than that of photoautotrophic cultivation (Laraib et al., 2021). Pang et al. introduced sodium gluconate using as an unconventional organic carbon source for the mixotrophic cultivation of *Haematococcus pluvialis* (Pang et al., 2019b). Yang et al. utilized sodium acetate to counter the growth inhibition of *Chlamydomonas reinhardtii* by nutrient deficiency, the lipid production increased by 93% under same nutrient deficiency condition (Yang et al., 2018). Li et al. reported that the lipid contents of *Chlorella* sp. strains were significantly higher under 1–30% CO_2 concentrations after adaptive evolution than those of the original strain (Li D. J. et al., 2015). Taking together, regardless of photoautotrophic, heterotrophic, or mixotrophic cultivation, optimum carbon concentrations are required to enhance the lipid accumulation.

Nitrogen source

Nitrogen is one of the nutrients most directly influencing the microalgal lipid accumulation (Shin et al., 2018; Feng et al., 2020). It may be provided in different forms, such as nitrate, urea or ammonium. However, different forms of nitrogen sources are specific for the growth and the lipid accumulation of microalgae. (Li et al., 2008; Gonzalez-Garcinuno et al., 2014; Feng et al., 2020). Of all the optimizing strategies of nitrogen source, nitrogen deficiency stress appears to be the most effective approach to stimulate the lipid accumulation in microalgae (Wu et al., 2013; Gonzalez-Garcinuno et al., 2014; Feng et al., 2020). For instance, Arguelles and Martinez-Goss found that there was an increasing trend in the lipid production of both of *Chlorolobion* sp. and *Chlorella* sp. under nitrogen limited culture condition in the range from 1.5 g/l to 0.375 g/l $NaNO_3$ (Arguelles and Martinez-Goss, 2021). The lipid productivity of *Chlorolobion* sp. and *Chlorella* sp. reached 227.84 and 151.14 mg/l/d at 0.375 g/l $NaNO_3$ concentration condition, respectively, 2.9-fold and 1.9-fold higher compared to at 1.5 g/l $NaNO_3$ growth condition (Arguelles and Martinez-Goss, 2021). Feng et al. demonstrated that *Chlorella* cells showed a 48.65% lipid content of dry cell weight under nitrogen limitation condition, 62% higher than that under nitrogen sufficient condition (Feng et al., 2020). Gao et al. investigated the effects of three types of nutrients (nitrogen, phosphate and iron)

starvation stress on the lipid production of *Chaetoceros muelleri* and *Dunaliella salina*, and observed the highest lipid contents when the cells cultured under a nitrogen deprivation condition (Gao et al., 2013).

It is widely acknowledged that by nitrogen deficiency stress, microalgae cell division decrease, channeling the lipid biosynthesis pathway toward neutral lipids rather than membrane lipids (Goncalves et al., 2016; Feng et al., 2020). As such, nitrogen limitation may also cause the production of biomass to decrease, which affects the further accumulation of lipids. A combination strategy of nitrogen deficiency stress and carbon enrichment might be an alternative option, supplying carbon source for fatty acid production along with stimulating the lipid biosynthesis pathway. Bharte and Desai revealed that the lipid contents of *Chlorella minutissima* and *Chlorella pyrenoidosa* were increased to 24 and 23% under the condition of nitrogen deprivation combined with acetate as additional carbon source, respectively 4 and 5.4% higher than those of a single nitrogen deprivation condition (Bharte and Desai, 2019). Zhu and Huang also reported that high glucose combined with low nitrogen could increase the lipid content of *Chlorella sorokiniana* under heterotrophic cultivation condition (Zhu and Huang, 2017).

Optimization of environmental conditions

The production of biomass and the accumulation of lipid of microalgae are coordinated with both nutrient conditions and environmental conditions (Rehman et al., 2022). Optimizing environmental parameters and regulating the environmental conditions to improve the synthesis of microalgal lipid is an efficient strategy to increase the microalgal lipid productivity and reduce the cost of microalgal biofuels production (Figure 2).

Temperature

Temperature is one of the most essential environmental parameters during the microalgal lipid production process, different temperatures not only change the nature of the nutrients but also affect the activities of various key enzymes in the metabolic process of the microalgal cells. Therefore, the effect of temperature on microalgal lipid accumulation is the comprehensive performance of various factors, and strain specific. In Wu et al.'s study, a decrease from 32.9 to 29.6% in the lipid content of a *Monoraphidium* strain was observed when the environmental temperature increased from 25 to 35°C (Wu et al., 2013). While in Converti et al.'s study, an increase in the temperature from 25 to 30°C led to a decrease in the lipid content of *C. vulgaris* from 14.71 to 5.90%, in contrast to that, the lipid content of *Nannochloropsis oculata* increased from 7.90 to 14.92% when the temperature increased from 25 to 30°C (Converti et al., 2009). Temperature plays more significant role in the growth of microalgae, however, to maximize lipid production of microalgae, it is essential to obtain robust and sufficient biomass as the foundation before the stage of lipid accumulation.

pH

Environmental pH value is a very significant comprehensive indicator of microalgal metabolic activities under certain circumstances, and has influences on the dynamic forms and the relative concentration of inorganic carbon source in the culture medium (Azov, 1982). Therefore, pH plays an important role in the cell growth and lipid accumulation of microalgae. Moheimani investigated the effects of pH value on the lipid productivities of *Tetraselmis suecica* and *Chlorella* sp. (Moheimani, 2013). The maximum lipid productivity of *T. suecica* reached 92 ± 13.1 mg/l/d when the pH value was kept at 7.5, and that of *Chlorella* sp. 99 ± 17.2 mg/l/d when the pH was kept at 7.0 (Moheimani, 2013). According to Zhang et al. the maximal lipid production (167.5 mg/l) of *Chlorella* sp. was also observed at initial pH of 7.0 (Zhang et al., 2014). Qiu et al. evaluated the effects of different pH on the lipid production of a strain of *Chlorella sorokiniana* (Qiu et al., 2017). By adjusting the pH value through feeding CO_2, an optimal pH for lipid accumulation was observed at 6.0, and the cetane numbers of biodiesel produced at pH 6.5, 7.0 and 7.5 fulfill the diesel standard (Qiu et al., 2017).

Light

Microalgae utilize light as driving force to obtain chemical energy during photoautotrophic cultivation (Blair et al., 2014). The availability of light, which could be achieved through multiple manipulations, is essential for cell growth and lipid accumulation of microalgae. For instance, Liu et al. investigated the effects of different light intensities on the lipid accumulation of *Scenedesmus* sp. and all the maximum biomass production, lipid content and neutral lipid content were obtained when the strain of *Scenedesmus* was cultivated at 400 μmol photons m^{-2} s^{-1} light intensity and limited nitrogen content (Liu et al., 2012). He et al. investigated the effects of different fluctuating light intensities on the lipid productivity of oleaginous microalgae and observed the maximum lipid productivities and neutral lipid contents when the microalgae were cultured under high fluctuating light intensity (He et al., 2015). Jung et al. found that green LED produced the highest lipid content while blue LED led to the highest biomass (Jung et al., 2019). By contrast, Sánchez-Saavedra et al. found that the lipid content was much higher under blue light under exponential growth phase compared with white, green and yellow light (Sanchez-Saavedra et al., 2020). Feng et al. investigated the effects of different light paths in a photobioreactor on the microalgal lipid content (Feng et al., 2020). Compared with higher light paths, the maximum lipid content was achieved when a strain of *Chlorella* was cultured at 5 cm light path of the flat plate photobioreactor (Feng et al., 2020). Lima et al. found that low-frequency flashing light can improve the fatty acid productivity

up to 3 times compared to that of the continuous light (Lima et al., 2021). Recently, nanoparticles have been emerging as spotlight for enhancing the microalgal lipid production by boosting the light conversion (Khanra et al., 2020; Vargas-Estrada et al., 2020).

Optimization of cultivation system

Staged cultivation

Cell growth of microalgae provides a solid foundation for the efficient biosynthesis of microalgal lipid. Optimization and regulation of nutritional and environmental conditions should not only supply superior conditions for cell growth but also satisfy the stress demand for lipid accumulation. In this sense, staged cultivation system emerged as a potential culture method applied for microalgal biofuel production since it is able to overcome the trade-off between the lipid accumulation with cell growth. In the staged cultivation system, commonly a two-stage approach, a nutrient-rich culture condition is provided in the first stage for maximum biomass, whereas stress culture conditions are applied for improving the microalgal lipid accumulation in the second stage (Figure 2).

Nutrient starvation, especially nitrogen starvation, has been regarded as the most reliable stress condition for stimulating the lipid production of microalgae in the staged cultivation system. Rai et al. employed nitrogen starvation condition during the two-stage cultivation of *Chlorella* sp. and the neutral lipid content increased to 50.43% of dry cell weight after 5-days of nitrogen starvation stress (Rai et al., 2017). Ghidossi et al. applied higher carbon to nitrogen ratio as nitrogen starvation stress condition in the second stage during two-stage culture process of *Chlorella protothecoides* and the lipid content reached 58% of dry cell weight with a lipid productivity of 2–5 fold higher compared with previously reported (Ghidossi et al., 2017). Other than nitrogen starvation, staged cultivation systems based on phosphorus starvation or silica starvation may also enhance microalgal lipid accumulation (Alvarez-Diaz et al., 2014; Smith et al., 2016). However, the effects of these nutrients starvation on the lipid accumulation seem to be more species dependent.

Considering the extra cost created in the transformation of microalgal biomass from nutrient-rich medium into nutrient-limitation medium when the nutrient starvation strategy is adopted in the lipid accumulation stage, nutrient starvation based staged cultivation system is less economically viable for the commercial application. In this regard, staged cultivation system based on environmental stress condition might be more feasible applied in large scale for microalgal lipid production. Ra et al. carried out a two-stage cultivation based on light stress condition and observed a maximal lipid content of 56% of dry cell weight in *Nannochloropsis oculata* under green light stress for 2 days in the second culture stage (Ra et al., 2016). By direct feeding of NaCl into the culture medium of *Scenedesmus obtusus* in the second stage, the lipid contents obtained under different concentrations of NaCl were all higher than that of single stage cultivation without the saline stress condition (Xia et al., 2013). A maximal lipid content of 47.7% of dry cell weight was obtained after 8 days stress of 20 g/l NaCl. When the same saline stress strategy applying at larger scale of 140 l, the lipid content of *S. obtusus* reached 42.1% of dry cell weight (Xia et al., 2013), illustrating that the staged cultivation system based on saline stress is not only effective but also practical for enhancing the microalgal lipid production. Higher salinity stress condition will induce higher lipid production, however, the tolerance of microalgae to high salinity is finite, and to high salinity could bring the growth inhibition even in the stage of lipid accumulation. To alleviate the biomass inhibition by high salinity stress for further enhancing the lipid accumulation, strategies for obtaining improved tolerance performance of microalgae to higher salinity stress have been proposed. For instance, Ho et al. administered a salinity gradient strategy by a stepwise sea salt addition during the two-stage cultivation of marine *Chlamydomonas*. The maximum lipid content reached 59.4% of dry cell weight after the optimal salinity gradient mode operated for 5 days, when 95% of nitrogen was consumed (Ho et al., 2014). In addition, the lipid productivity was also much higher compared with that of other staged cultivation system (Ho et al., 2014). This synergistic operation combining salinity gradient stress with nitrogen starvation stress develops positive impact on the trade-off between lipid production with lipid productivity during the staged cultivation process of microalgae. Furthermore, for more complex staged culture system based on multi-stress condition, Wang et al. integrated glucose fed-batch operation in the cell growth stage and hyperosmotic combined with nitrogen starvation stress condition in the lipid accumulation stage during the two-stage cultivation process of *Chlorella protothecoides* (Wang et al., 2017). The lipid content, lipid yield on glucose and lipid productivity obtained in the multi-stress two-stage fed-batch culture system increased 1.92, 1.79 and 1.60-fold compared to a no stress single stage fed-batch culture system, respectively (Wang et al., 2017).

Applying staged cultivation system based on nitrogen starvation stress is one of the most reliable strategies for enhancing the microalgal lipid production. While environmental stress operation is practically more feasible utilizing in large scale for industrialization. Consequently, integrating environmental stress operation when the culture system of microalgae is under nitrogen starvation stress condition may significantly enhance the microalgal lipid production during the large-scale staged cultivation process. The optimization and regulation of the entire stress-integrated staged cultivation process are of great importance for the commercialization of microalgal lipid production.

Heterotrophic and mixotrophic cultivation

Some species of microalgae have the ability to grow heterotrophically or mixotrophically, displaying considerable metabolic diversity and flexibility (Morales-Sanchez et al., 2017; Pang et al., 2019a). Heterotrophic cultivation refers to

the culture mode in which microalgae exclusively use organic substrates as both carbon and energy source (Yin et al., 2020). Liu et al. compared the effects of culture modes on the cell growth and lipid yield of *Chlorella* sp., heterotrophic culture mode was regarded as the optimal strategy for the accumulation of microalgal lipid (Liu et al., 2019). Ghidossi et al. investigated the effect of C/N ratio on the microalgal growth and lipid productivity of *Chlorella protothecoides* under heterotrophic cultivation, the maximum lipid productivity and microalgal biomass reached 16.7 g/l/d and 255 g/l, respectively, when the different C/N ratios were combined during the heterotrophic cultivation process (Ghidossi et al., 2017). Heterotrophic cultivation mode could easily manipulate the C/N ratio for optimizing the microalgal cell growth and the lipid accumulation. However, it worth noting that the economic viability of the heterotrophic cultivation utilizing the organic substrates.

Mixotrophic cultivation is a special mode in which microalgae can metabolize both organic and inorganic carbon source simultaneously under solar energy (Wang et al., 2014). Mixotrophic cultivation may be an ideal culture mode for large-scale microalgal lipid production because of its combined advantages of synergism of photoautotrophic and heterotrophic cultivation (Pang et al., 2019a; Patel et al., 2020). Heredia-Arroyo et al. studied the mixotrophic cultivations of *Chlorella vulgaris* by different organic carbon source at different concentrations (Heredia-Arroyo et al., 2011). 0.19 g/l of lipid production was obtained when 4 g/l glucose was applied as the initial organic carbon source, 77% higher than that of photoautotrophic culture condition (Heredia-Arroyo et al., 2011). Gao et al. regulated the ratio of organic carbon and nitrogen source during the mixotrophic cultivation of *Chlorella* sp. G-9 and observed a 36.5% of lipid content of dry cell weight and a lipid productivity of 32.6 mg/l/d, 13-fold higher than that in photoautotrophic culture condition (Gao et al., 2019). Based on the mixotrophic culture mode, Xue et al. constructed a co-culture system of *Spirulina platensis* and yeast *Rhodotorula glutinis* (Xue et al., 2010). Four hundred sixty-seven milligram per liter of total lipid production was obtained, 2-fold higher than the sum of that of the two single systems under the same mixotrophic culture condition (Xue et al., 2010). Qin et al. constructed a co-culture system of *Chlorella pyrenoidosa* and yeast *Yarrowia lipolytica* using glycerol as organic carbon source during the mixotrophic cultivation (Qin et al., 2019). The maximum lipid production reached 0.77 g/l, 2.85-fold higher than that of single system of *Y. lipolytica* and 3.53-fold higher than that of the single system of *C. pyrenoidosa* (Qin et al., 2019).

Compared with the improved lipid production or productivity, mixotrophic cultivation has a much more pronounced influence on the microalgae cell growth. Consequently, the enhancement of lipid content of microalgae seems to be less obvious or barely noticeable. It might be favorable to integrate the mixotrophic cultivation mode with staged cultivation system based on stress condition to accumulate higher microalgal lipid with good performance in terms of overcoming the trade-off between the lipid accumulation with the cell growth.

TABLE 1 Selected species of microalgae accumulate lipid for biofuel production.

Microalgal species	Lipid content	References
Phaeodactylum tricornutum	60.6% dry cell weight (DCW)	Jung et al. (2019)
Chlamydomonas sp.	59.4% of DCW	Ho et al. (2014)
Chlorella protothecoides	58% of DCW	Ghidossi et al. (2017)
Phaeodactylum tricornutum	57.5% of DCW	Zou et al. (2018)
Nannochloropsis oculata	56% of DCW	Ra et al. (2016)
Phaeodactylum tricornutum	55.7% of DCW	Xue et al. (2017)
Chlorella sp.	53.5% of DCW	Feng et al. (2020)
Scenedesmus obliquus	49.4% of DCW	de Jaeger et al. (2014)
Scenedesmus obtusus	47.7% of DCW	Xia et al. (2013)
Nannochloropsis oceanica	42.9% of DCW	Chen et al. (2017)
Chlorella vulgaris	39% of DCW	Laraib et al. (2021)
Chlorella sorokiniana	32% of DCW	Zhu and Huang (2017)

Concluding remarks and future perspectives

Microalgae have been drawing tremendous attention as a promising emerging feedstock for the production of lipid-based biofuels. Economical and commercial application of microalgal biofuel production is subject to the enhancement of lipid accumulation on the basis of overcoming the conflicts between microalgal cell growth and lipid accumulation. Extensive efforts have been made on improving microalgal lipid accumulation including genetic modifications of microalgal strains by metabolic engineering and process regulations of microalgae cultivation by integrating multiple optimization strategies widely applied in industrial microbiology (Table 1). In future, in-depth understanding of the microalgal lipid metabolic network is essential for the construction of high-performance microalgal strains through metabolic engineering and molecular modification. Emerging omics techniques, including metabolomics, proteomics, and lipidomics, have been exhibiting great potential for further identifying and understanding of the microalgal lipid biosynthetic pathways by cooperating with genetic engineering (Arora et al., 2018; Rawat et al., 2021). Systematic optimization strategies integrating various biomass improvement strategies with nutrient and environmental stress operation during the staged cultivation mode should be developed for the maximization of microalgal lipid accumulation. More assessment of these synergistic strategies applying in large-scale microalgal lipid production with economic feasibility are still required. These advancements for enhancing microalgal lipid accumulation are certainly making biofuel production based on microalgae a reality for commercial application in the near future.

Author contributions

ZZ and XL conceived the outline and drafted the manuscript. JS and YF revised the manuscript. PL made major revisions of the manuscript. All authors contributed to the article and approved the submitted version.

Funding

This research was funded by Open Project Funding of the State Key Laboratory of Biocatalysis and Enzyme Engineering (Grant Number SKLBEE2020015), CAS Key Laboratory of Bio-based Materials, Chinese Academy of Sciences (Grant Number BMF-2020-08), Natural Science Foundation of the Jiangsu Higher Education Institutions of China (Grant Number 20KJB180012), and the China Scholarship Council (Scholarship Number 202108320126).

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

References

Alvarez-Diaz, P. D., Ruiz, J., Arbib, Z., Barragan, J., Garrido-Perez, C., and Perales, J. A. (2014). Lipid production of microalga *Ankistrodesmus falcatus* increased by nutrient and light starvation in a two-stage cultivation process. *Appl. Biochem. Biotechnol.* 174, 1471–1483. doi: 10.1007/s12010-014-1126-5

Amorim, M. L., Soares, J., Coimbra, J. S. D., Leite, M. D., Albino, L. F. T., and Martins, M. A. (2021). Microalgae proteins: production, separation, isolation, quantification, and application in food and feed. *Crit. Rev. Food Sci. Nutr.* 61, 1976–2002. doi: 10.1080/10408398.2020.1768046

Anto, S., Mukherjee, S. S., Muthappa, R., Mathimani, T., Deviram, G., Kumar, S. S., et al. (2020). Algae as green energy reserve: technological outlook on biofuel production. *Chemosphere* 242:125079. doi: 10.1016/j.chemosphere.2019.125079

Arguelles, E. D., and Martinez-Goss, M. R. (2021). Lipid accumulation and profiling in microalgae *Chlorolobion* sp. (BIOTECH 4031) and *chlorella* sp. (BIOTECH 4026) during nitrogen starvation for biodiesel production. *J. Appl. Phycol.* 33, 1–11. doi: 10.1007/s10811-020-02126-z

Arora, N., Pienkos, P. T., Pruthi, V., Poluri, K. M., and Guarnieri, M. T. (2018). Leveraging algal omics to reveal potential targets for augmenting TAG accumulation. *Biotechnol. Adv.* 36, 1274–1292. doi: 10.1016/j.biotechadv.2018.04.005

Azov, Y. (1982). Effect of pH on inorganic carbon uptake in algal cultures. *Appl. Environ. Microbiol.* 43, 1300–1306. doi: 10.1128/aem.43.6.1300-1306.1982

Behera, B., Unpaprom, Y., Ramaraj, R., Maniam, G. P., Govindan, N., and Paramasivan, B. (2021). Integrated biomolecular and bioprocess engineering strategies for enhancing the lipid yield from microalgae. *Renew. Sust. Energ. Rev.* 148:111270. doi: 10.1016/j.rser.2021.111270

Bharte, S., and Desai, K. (2019). The enhanced lipid productivity of *Chlorella minutissima* and *Chlorella pyrenoidosa* by carbon coupling nitrogen manipulation for biodiesel production. *Environ. Sci. Pollut. R.* 26, 3492–3500. doi: 10.1007/s11356-018-3757-5

Blair, M. F., Kokabian, B., and Gude, V. G. (2014). Light and growth medium effect on *Chlorella vulgaris* biomass production. *J. Environ. Chem. Eng.* 2, 665–674. doi: 10.1016/j.jece.2013.11.005

Brar, A., Kumar, M., Soni, T., Vivekanand, V., and Pareek, N. (2021). Insights into the genetic and metabolic engineering approaches to enhance the competence of microalgae as biofuel resource: a review. *Bioresour. Technol.* 339:125597. doi: 10.1016/j.biortech.2021.125597

Chen, C. Y., Kao, A. L., Tsai, Z. C., Chow, T. J., Chang, H. Y., Zhao, X. Q., et al. (2016). Expression of type 2 diacylglycerol acyltransferse gene DGTT1 from *Chlamydomonas reinhardtii* enhances lipid production in *Scenedesmus obliquus*. *Biotechnol. J.* 11, 336–344. doi: 10.1002/biot.201500272

Chen, J. W., Liu, W. J., Hu, D. X., Wang, X., Balamurugan, S., Alimujiang, A., et al. (2017). Identification of a malonyl CoA-acyl carrier protein transacylase and its regulatory role in fatty acid biosynthesis in oleaginous microalga *Nannochloropsis oceanica*. *Biotechnol. Appl. Biochem.* 64, 620–626. doi: 10.1002/bab.1531

Chisti, Y. (2007). Biodiesel from microalgae. *Biotechnol. Adv.* 25, 294–306. doi: 10.1016/j.biotechadv.2007.02.001

Chu, W. L. (2017). Strategies to enhance production of microalgal biomass and lipids for biofuel feedstock. *Eur. J. Phycol.* 52, 419–437. doi: 10.1080/09670262.2017.1379100

Converti, A., Casazza, A. A., Ortiz, E. Y., Perego, P., and Del Borghi, M. (2009). Effect of temperature and nitrogen concentration on the growth and lipid content of *Nannochloropsis oculata* and *Chlorella vulgaris* for biodiesel production. *Chem. Eng. Process.* 48, 1146–1151. doi: 10.1016/j.cep.2009.03.006

Das, M., Patra, P., and Ghosh, A. (2020). Metabolic engineering for enhancing microbial biosynthesis of advanced biofuels. *Renew. Sust. Energ. Rev.* 119:109562. doi: 10.1016/j.rser.2019.109562

de Jaeger, L., Verbeek, R. E. M., Draaisma, R. B., Martens, D. E., Springer, J., Eggink, G., et al. (2014). Superior triacylglycerol (TAG) accumulation in starchless mutants of *Scenedesmus obliquus*: (I) mutant generation and characterization. *Biotechnol. Biofuels* 7:69. doi: 10.1186/1754-6834-7-69

Du, G., Liu, L., Li, Y., and Chen, J. (2006). Optimization of fermentation process for achieving high product concentration, high yield and high productivity. *Chem. Ind. Eng. Prog.* 25, 1128–1133.

Feng, P. Z., Xu, Z. B., Qin, L., Alam, M. A., Wang, Z. M., and Zhu, S. N. (2020). Effects of different nitrogen sources and light paths of flat plate photobioreactors on the growth and lipid accumulation of *chlorella* sp. GN1 outdoors. *Bioresour. Technol.* 301:122762. doi: 10.1016/j.biortech.2020.122762

Gao, F., Yang, H. L., Li, C., Peng, Y. Y., Lu, M. M., Jin, W. H., et al. (2019). Effect of organic carbon to nitrogen ratio in wastewater on growth, nutrient uptake and lipid accumulation of a mixotrophic microalgae *chlorella* sp. *Bioresour. Technol.* 282, 118–124. doi: 10.1016/j.biortech.2019.03.011

Gao, Y., Yang, M., and Wang, C. (2013). Nutrient deprivation enhances lipid content in marine microalgae. *Bioresour. Technol.* 147, 484–491. doi: 10.1016/j.biortech.2013.08.066

Ghidossi, T., Marison, I., Devery, R., Gaffney, D., and Forde, C. (2017). Characterization and optimization of a fermentation process for the production of high cell densities and lipids using heterotrophic cultivation of *Chlorella protothecoides*. *Ind. Biotechnol.* 13, 253–259. doi: 10.1089/ind.2017.0007

Gomma, A. E., Lee, S. K., Sun, S. M., Yang, S. H., and Chung, G. (2015). Improvement in oil production by increasing malonyl-CoA and glycerol-3-phosphate pools in *Scenedesmus quadricauda*. *Indian J. Microbiol.* 55, 447–455. doi: 10.1007/s12088-015-0546-4

Goncalves, E. C., Wilkie, A. C., Kirst, M., and Rathinasabapathi, B. (2016). Metabolic regulation of triacylglycerol accumulation in the green algae: identification of potential targets for engineering to improve oil yield. *Plant Biotechnol. J.* 14, 1649–1660. doi: 10.1111/pbi.12523

Gonzalez-Garcinuno, A., Tabernero, A., Sanchez-Alvarez, J. M., del Valle, E. M. M., and Galan, M. A. (2014). Effect of nitrogen source on growth and lipid accumulation in *Scenedesmus abundans* and *Chlorella ellipsoidea*. *Bioresour. Technol.* 173, 334–341. doi: 10.1016/j.biortech.2014.09.038

Harun, R., Danquah, M. K., and Forde, G. M. (2010). Microalgal biomass as a fermentation feedstock for bioethanol production. *J. Chem. Technol. Biotechnol.* 85:203. doi: 10.1002/jctb.2287

Haslam, R. P., Hamilton, M. L., Economou, C. K., Smith, R., Hassall, K. L., Napier, J. A., et al. (2020). Overexpression of an endogenous type 2 diacylglycerol acyltransferase in the marine diatom *Phaeodactylum tricornutum* enhances lipid production and omega-3 long-chain polyunsaturated fatty acid content. *Biotechnol. Biofuels* 13:87. doi: 10.1186/s13068-020-01726-8

He, Q. N., Yang, H. J., Xu, L. L., Xia, L., and Hu, C. X. (2015). Sufficient utilization of natural fluctuating light intensity is an effective approach of promoting lipid productivity in oleaginous microalgal cultivation outdoors. *Bioresour. Technol.* 180, 79–87. doi: 10.1016/j.biortech.2014.12.088

Heredia-Arroyo, T., Wei, W., Ruan, R., and Hu, B. (2011). Mixotrophic cultivation of *Chlorella vulgaris* and its potential application for the oil accumulation from non-sugar materials. *Biomass Bioenergy* 35, 2245–2253. doi: 10.1016/j.biombioe.2011.02.036

Ho, S. H., Nakanishi, A., Ye, X., Chang, J. S., Hara, K., Hasunuma, T., et al. (2014). Optimizing biodiesel production in marine *Chlamydomonas* sp. JSC4 through metabolic profiling and an innovative salinity-gradient strategy. *Biotechnol. Biofuels* 7:97. doi: 10.1186/1754-6834-7-97

Hsieh, H. J., Su, C. H., and Chien, L. J. (2012). Accumulation of lipid production in *Chlorella minutissima* by triacylglycerol biosynthesis-related genes cloned from *Saccharomyces cerevisiae* and *Yarrowia lipolytica*. *J. Microbiol.* 50, 526–534. doi: 10.1007/s12275-012-2041-5

Jung, J. H., Sirisuk, P., Ra, C. H., Kim, J. M., Jeong, G. T., and Kim, S. K. (2019). Effects of green LED light and three stresses on biomass and lipid accumulation with two-phase culture of microalgae. *Process Biochem.* 77, 93–99. doi: 10.1016/j.procbio.2018.11.014

Kao, P. H., and Ng, I. S. (2017). CRISPRi mediated phosphoenolpyruvate carboxylase regulation to enhance the production of lipid in *Chlamydomonas reinhardtii*. *Bioresour. Technol.* 245, 1527–1537. doi: 10.1016/j.biortech.2017.04.111

Khan, S., and Fu, P. C. (2020). Biotechnological perspectives on algae: a viable option for next generation biofuels. *Curr. Opin. Biotechnol.* 62, 146–152. doi: 10.1016/j.copbio.2019.09.020

Khanra, A., Vasistha, S., Kumar, P., and Rai, M. P. (2020). Role of C/N ratio on microalgae growth in mixotrophy and incorporation of titanium nanoparticles for cell flocculation and lipid enhancement in economical biodiesel application. *Biotech* 10:331. doi: 10.1007/s13205-020-02323-0

Korkhovoy, V. I., and Blume, Y. B. (2013). Biodiesel from microalgae: ways for increasing the effectiveness of lipid accumulation by genetic engineering methods. *Cytol. Genet.* 47, 349–358. doi: 10.3103/S0095452713060030

Laraib, N., Manzoor, M., Javid, A., Jabeen, F., Bukhari, S. M., Ali, W., et al. (2021). Mixotrophic cultivation of *Chlorella vulgaris* in sugarcane molasses preceding nitrogen starvation: biomass productivity, lipid content, and fatty acid analyses. *Environ. Prog. Sustain. Energy* 40:e13625. doi: 10.1002/ep.13625

Lei, A. P., Chen, H., Shen, G. M., Hu, Z. L., Chen, L., and Wang, J. X. (2012). Expression of fatty acid synthesis genes and fatty acid accumulation in *Haematococcus pluvialis* under different stressors. *Biotechnol. Biofuels* 5:18. doi: 10.1186/1754-6834-5-18

Li, T. T., Gargouri, M., Feng, J., Park, J. J., Gao, D. F., Miao, C., et al. (2015). Regulation of starch and lipid accumulation in a microalga *Chlorella sorokiniana*. *Bioresour. Technol.* 180, 250–257. doi: 10.1016/j.biortech.2015.01.005

Li, Y. T., Han, D. X., Hu, G. R., Dauvillee, D., Sommerfeld, M., Ball, S., et al. (2010). *Chlamydomonas* starchless mutant defective in ADP-glucose pyrophosphorylase hyper-accumulates triacylglycerol. *Metab. Eng.* 12, 387–391. doi: 10.1016/j.ymben.2010.02.002

Li, Y. Q., Horsman, M., Wang, B., Wu, N., and Lan, C. Q. (2008). Effects of nitrogen sources on cell growth and lipid accumulation of green alga *Neochloris oleoabundans*. *Appl. Microbiol. Biotechnol.* 81, 629–636. doi: 10.1007/s00253-008-1681-1

Li, Z. P., Meng, T., Ling, X. P., Li, J., Zheng, C. Q., Shi, Y. Y., et al. (2018). Overexpression of malonyl-CoA: ACP Transacylase in *Schizochytrium* sp. to improve polyunsaturated fatty acid production. *J. Agric. Food Chem.* 66, 5382–5391. doi: 10.1021/acs.jafc.8b01026

Li, D. J., Wang, L., Zhao, Q. Y., Wei, W., and Sun, Y. H. (2015). Improving high carbon dioxide tolerance and carbon dioxide fixation capability of *chlorella* sp by adaptive laboratory evolution. *Bioresour. Technol.* 185, 269–275. doi: 10.1016/j.biortech.2015.03.011

Li, T. T., Zheng, Y. B., Yu, L., and Chen, S. L. (2014). Mixotrophic cultivation of a *Chlorella sorokiniana* strain for enhanced biomass and lipid production. *Biomass Bioenergy* 66, 204–213. doi: 10.1016/j.biombioe.2014.04.010

Lima, S., Schulze, P. S. C., Schuler, L. M., Rautenberger, R., Morales-Sanchez, D., Santos, T. F., et al. (2021). Flashing light emitting diodes (LEDs) induce proteins, polyunsaturated fatty acids and pigments in three microalgae. *J. Biotechnol.* 325, 15–24. doi: 10.1016/j.jbiotec.2020.11.019

Liu, X., Hong, Y., Liu, P., Zhan, J., and Yan, R. (2019). Effects of cultivation strategies on the cultivation of *chlorella* sp. HQ in photoreactors. *Front. Environ. Sci. Eng.* 13:78. doi: 10.1007/s11783-019-1162-z

Liu, J. H., Yuan, C., Hu, G. R., and Li, F. L. (2012). Effects of light intensity on the growth and lipid accumulation of microalga *Scenedesmus* sp. 11-1 under nitrogen limitation. *Appl. Biochem. Biotechnol.* 166, 2127–2137. doi: 10.1007/s12010-012-9639-2

Markou, G., and Nerantzis, E. (2013). Microalgae for high-value compounds and biofuels production: a review with focus on cultivation under stress conditions. *Biotechnol. Adv.* 31, 1532–1542. doi: 10.1016/j.biotechadv.2013.07.011

Markou, G., Vandamme, D., and Muylaert, K. (2014). Microalgal and cyanobacterial cultivation: the supply of nutrients. *Water Res.* 65, 186–202. doi: 10.1016/j.watres.2014.07.025

Medipally, S. R., Yusoff, F. M., Banerjee, S., and Shariff, M. (2015). Microalgae as sustainable renewable energy feedstock for biofuel production. *Biomed. Res. Int.* 2015:519513. doi: 10.1155/2015/519513

Moffett, J. R., Puthillathu, N., Vengilote, R., Jaworski, D. M., and Namboodiri, A. M. (2020). Acetate revisited: a key biomolecule at the nexus of metabolism, epigenetics and oncogenesis-part 1: acetyl-CoA, Acetogenesis and acyl-CoA short-chain Synthetases. *Front. Physiol.* 11:580167. doi: 10.3389/fphys.2020.580167

Moheimani, N. R. (2013). Inorganic carbon and pH effect on growth and lipid productivity of *Tetraselmis suecica* and *chlorella* sp. (*Chlorophyta*) grown outdoors in bag photobioreactors. *J. Appl. Phycol.* 25, 387–398. doi: 10.1007/s10811-012-9873-6

Morales-Sanchez, D., Martinez-Rodriguez, O. A., and Martinez, A. (2017). Heterotrophic cultivation of microalgae: production of metabolites of commercial interest. *J. Chem. Technol. Biotechnol.* 92, 925–936. doi: 10.1002/jctb.5115

Mulgund, A. (2022). Increasing lipid accumulation in microalgae through environmental manipulation, metabolic and genetic engineering: a review in the energy NEXUS framework. *Energy Nexus* 5:100054. doi: 10.1016/j.nexus.2022.100054

Niu, Y. F., Wang, X., Hu, D. X., Balamurugan, S., Li, D. W., Yang, W. D., et al. (2016). Molecular characterization of a glycerol-3-phosphate acyltransferase reveals key features essential for triacylglycerol production in *Phaeodactylum tricornutum*. *Biotechnol. Biofuels* 9:60. doi: 10.1186/s13068-016-0478-1

Norashikin, M. N., Loh, S. H., Aziz, A., and Cha, T. S. (2018). Metabolic engineering of fatty acid biosynthesis in *Chlorella vulgaris* using an endogenous omega-3 fatty acid desaturase gene with its promoter. *Algal Res.* 31, 262–275. doi: 10.1016/j.algal.2018.02.020

Ong, H. C., Chen, W.-H., Singh, Y., Gan, Y. Y., Chen, C.-Y., and Show, P. L. (2020). A state-of-the-art review on thermochemical conversion of biomass for biofuel production: a TG-FTIR approach. *Energy Convers. Manag.* 209:112634. doi: 10.1016/j.enconman.2020.112634

Pang, N., Gu, X. Y., Chen, S. L., Kirchhoff, H., Lei, H. W., and Roje, S. (2019a). Exploiting mixotrophy for improving productivities of biomass and co-products of microalgae. *Renew. Sust. Energ. Rev.* 112, 450–460. doi: 10.1016/j.rser.2019.06.001

Pang, N., Gu, X. Y., Fu, X., and Chen, S. L. (2019b). Effects of gluconate on biomass improvement and light stress tolerance of *Haematococcus pluvialis* in mixotrophic culture. *Algal Res.* 43:101647. doi: 10.1016/j.algal.2019.101647

Patel, A. K., Choi, Y. Y., and Sim, S. J. (2020). Emerging prospects of mixotrophic microalgae: way forward to sustainable bioprocess for environmental remediation and cost-effective biofuels. *Bioresour. Technol.* 300:122741. doi: 10.1016/j.biortech.2020.122741

Peng, L., Fu, D., Chu, H., Wang, Z., and Qi, H. (2020). Biofuel production from microalgae: a review. *Environ. Chem. Lett.* 18, 285–297. doi: 10.1007/s10311-019-00939-0

Qin, L., Liu, L., Wang, Z. M., Chen, W. N., and Wei, D. (2019). The mixed culture of microalgae *Chlorella pyrenoidosa* and yeast *Yarrowia lipolytica* for microbial biomass production. *Bioprocess Biosyst. Eng.* 42, 1409–1419. doi: 10.1007/s00449-019-02138-1

Qiu, R., Gao, S., Lopez, P. A., and Ogden, K. L. (2017). Effects of pH on cell growth, lipid production and CO_2 addition of microalgae *Chlorella sorokiniana*. *Algal Res.* 28, 192–199. doi: 10.1016/j.algal.2017.11.004

Ra, C. H., Kang, C. H., Jung, J. H., Jeong, G. T., and Kim, S. K. (2016). Effects of light-emitting diodes (LEDs) on the accumulation of lipid content using a two-phase culture process with three microalgae. *Bioresour. Technol.* 212, 254–261. doi: 10.1016/j.biortech.2016.04.059

Rai, V., Muthuraj, M., Gandhi, M. N., Das, D., and Srivastava, S. (2017). Real-time iTRAQ-based proteome profiling revealed the central metabolism involved in nitrogen starvation induced lipid accumulation in microalgae. *Sci. Rep.* 7:45732. doi: 10.1038/srep45732

Ramirez-Lopez, C., Chairez, I., and Fernandez-Linares, L. (2016). A novel culture medium designed for the simultaneous enhancement of biomass and lipid production by *Chlorella vulgaris* UTEX 26. *Bioresour. Technol.* 212, 207–216. doi: 10.1016/j.biortech.2016.04.051

Ran, W. Y., Wang, H. T., Liu, Y. H., Qi, M., Xiang, Q., Yao, C. H., et al. (2019). Storage of starch and lipids in microalgae: biosynthesis and manipulation by nutrients. *Bioresour. Technol.* 291:121894. doi: 10.1016/j.biortech.2019.121894

Rawat, J., Gupta, P. K., Pandit, S., Prasad, R., and Pande, V. (2021). Current perspectives on integrated approaches to enhance lipid accumulation in microalgae. *Biotech* 11:303. doi: 10.1007/s13205-021-02851-3

Rehman, M., Kesharvani, S., Dwivedi, G., and Suneja, K. G. (2022). Impact of cultivation conditions on microalgae biomass productivity and lipid content. *Mat. Today Proc.* 56, 282–290. doi: 10.1016/j.matpr.2022.01.152

Sanchez-Saavedra, M. D. P., Sauceda-Carvajal, D., Castro-Ochoa, F. Y., and Molina-Cardenas, C. A. (2020). The use of light spectra to improve the growth and lipid content of *Chlorella vulgaris* for biofuels production. *Bioenergy Res.* 13, 487–498. doi: 10.1007/s12155-019-10070-1

Saranya, D., and Shanthakumar, S. (2021). Insights into the influence of CO_2 supplement on phycoremediation and lipid accumulation potential of microalgae: an exploration for biodiesel production. *Environ. Technol. Innov.* 23:101596. doi: 10.1016/j.eti.2021.101596

Shin, Y. S., Choi, H. I., Choi, J. W., Lee, J. S., Sung, Y. J., and Sim, S. J. (2018). Multilateral approach on enhancing economic viability of lipid production from microalgae: a review. *Bioresour. Technol.* 258, 335–344. doi: 10.1016/j.biortech.2018.03.002

Shokravi, Z., Shokravi, H., Chyuan, O. H., Lau, W. J., Koloor, S. S. R., Petru, M., et al. (2020). Improving 'lipid productivity' in microalgae by bilateral enhancement of biomass and lipid contents: a review. *Sustainability* 12:9083. doi: 10.3390/su12219083

Singh, P., Kumari, S., Guldhe, A., Misra, R., Rawat, I., and Bux, F. (2016). Trends and novel strategies for enhancing lipid accumulation and quality in microalgae. *Renew. Sust. Energ. Rev.* 55, 1–16. doi: 10.1016/j.rser.2015.11.001

Slade, R., and Bauen, A. (2013). Micro-algae cultivation for biofuels: cost, energy balance, environmental impacts and future prospects. *Biomass Bioenergy* 53, 29–38. doi: 10.1016/j.biombioe.2012.12.019

Smith, S. R., Gle, C., Abbriano, R. M., Traller, J. C., Davis, A., Trentacoste, E., et al. (2016). Transcript level coordination of carbon pathways during silicon starvation-induced lipid accumulation in the diatom *Thalassiosira pseudonana*. *New Phytol.* 210, 890–904. doi: 10.1111/nph.13843

Srivastava, R. K., Shetti, N. P., Reddy, K. R., and Aminabhavi, T. M. (2020). Biofuels, biodiesel and biohydrogen production using bioprocesses. *Environ. Chem. Lett.* 18, 1049–1072. doi: 10.1007/s10311-020-00999-7

Sun, X. M., Ren, L. J., Zhao, Q. Y., Ji, X. J., and Huang, H. (2019). Enhancement of lipid accumulation in microalgae by metabolic engineering. *BBA-Mol. Cell Biol. L.* 1864, 552–566. doi: 10.1016/j.bbalip.2018.10.004

Takeshita, T., Ota, S., Yamazaki, T., Hirata, A., Zachleder, V., and Kawano, S. (2014). Starch and lipid accumulation in eight strains of six *chlorella* species under comparatively high light intensity and aeration culture conditions. *Bioresour. Technol.* 158, 127–134. doi: 10.1016/j.biortech.2014.01.135

Tang, D. H., Han, W., Li, P. L., Miao, X. L., and Zhong, J. J. (2011). CO_2 biofixation and fatty acid composition of *Scenedesmus obliquus* and *Chlorella pyrenoidosa* in response to different CO_2 levels. *Bioresour. Technol.* 102, 3071–3076. doi: 10.1016/j.biortech.2010.10.047

Tian, J. H., Zheng, M. G., Yang, G. P., Zheng, L., Chen, J. L., and Yang, B. J. (2013). Cloning and stress-responding expression analysis of malonyl CoA-acyl carrier protein transacylase gene of *Nannochloropsis gaditana*. *Gene* 530, 33–38. doi: 10.1016/j.gene.2013.08.002

Trentacoste, E. M., Shrestha, R. P., Smith, S. R., Gle, C., Hartmann, A. C., Hildebrand, M., et al. (2013). Metabolic engineering of lipid catabolism increases microalgal lipid accumulation without compromising growth. *Proc. Natl. Acad. Sci. U. S. A.* 110, 19748–19753. doi: 10.1073/pnas.1309299110

Vargas-Estrada, L., Torres-Arellano, S., Longoria, A., Arias, D. M., Okoye, P. U., and Sebastiana, P. J. (2020). Role of nanoparticles on microalgal cultivation: a review. *Fuel* 280:118598. doi: 10.1016/j.fuel.2020.118598

Wang, L., Chen, L., Yang, S., and Tan, X. (2020). Photosynthetic conversion of carbon dioxide to oleochemicals by cyanobacteria: recent advances and future perspectives. *Front. Microbiol.* 11:634. doi: 10.3389/fmicb.2020.00634

Wang, X., Dong, H. P., Wei, W., Balamurugan, S., Yang, W. D., Liu, J. S., et al. (2018). Dual expression of plastidial GPAT1 and LPAT1 regulates triacylglycerol production and the fatty acid profile in *Phaeodactylum tricornutum*. *Biotechnol. Biofuels* 11:318. doi: 10.1186/s13068-018-1317-3

Wang, Q. T., Feng, Y. B., Lu, Y. D., Xin, Y., Shen, C., Wei, L., et al. (2021). Manipulating fatty-acid profile at unit chain-length resolution in the model industrial oleaginous microalgae *Nannochloropsis*. *Metab. Eng.* 66, 157–166. doi: 10.1016/j.ymben.2021.03.015

Wang, J. L., Singer, S. D., Souto, B. A., Asomaning, J., Ullah, A., Bressler, D. C., et al. (2022). Current progress in lipid-based biofuels: feedstocks and production technologies. *Bioresour. Technol.* 351:127020. doi: 10.1016/j.biortech.2022.127020

Wang, T., Tian, X., Liu, T., Wang, Z., Guan, W., Guo, M., et al. (2017). A two-stage fed-batch heterotrophic culture of *Chlorella protothecoides* that combined nitrogen depletion with hyperosmotic stress strategy enhanced lipid yield and productivity. *Process Biochem.* 60, 74–83. doi: 10.1016/j.procbio.2017.05.027

Wang, J. H., Yang, H. Z., and Wang, F. (2014). Mixotrophic cultivation of microalgae for biodiesel production: status and prospects. *Appl. Biochem. Biotechnol.* 172, 3307–3329. doi: 10.1007/s12010-014-0729-1

Wu, L. F., Chen, P. C., and Lee, C. M. (2013). The effects of nitrogen sources and temperature on cell growth and lipid accumulation of microalgae. *Int. Biodeterior. Biodegrad.* 85, 506–510. doi: 10.1016/j.ibiod.2013.05.016

Xia, L., Ge, H., Zhou, X., Zhang, D., and Hu, C. (2013). Photoautotrophic outdoor two-stage cultivation for oleaginous microalgae *Scenedesmus obtusus* XJ-15. *Bioresour. Technol.* 144, 261–267. doi: 10.1016/j.biortech.2013.06.112

Xue, J., Balamurugan, S., Li, D. W., Liu, Y. H., Zeng, H., Wang, L., et al. (2017). Glucose-6-phosphate dehydrogenase as a target for highly efficient fatty acid biosynthesis in microalgae by enhancing NADPH supply. *Metab. Eng.* 41, 212–221. doi: 10.1016/j.ymben.2017.04.008

Xue, F. Y., Miao, J. X., Zhang, X., and Tan, T. W. (2010). A new strategy for lipid production by mix cultivation of *Spirulina platensis* and *Rhodotorula glutinis*. *Appl. Biochem. Biotechnol.* 160, 498–503. doi: 10.1007/s12010-008-8376-z

Yang, L., Chen, J., Qin, S., Zeng, M., Jiang, Y. G., Hu, L., et al. (2018). Growth and lipid accumulation by different nutrients in the microalga Chlamydomonas reinhardtii. *Biotechnol. Biofuels* 11:40. doi: 10.1186/s13068-018-1041-z

Yin, Z. H., Zhu, L. D., Li, S. X., Hu, T. Y., Chu, R. Y., Mo, F., et al. (2020). A comprehensive review on cultivation and harvesting of microalgae for biodiesel production: environmental pollution control and future directions. *Bioresour. Technol.* 301:122804. doi: 10.1016/j.biortech.2020.122804

Zhang, Q., Wang, T., and Hong, Y. (2014). Investigation of initial pH effects on growth of an oleaginous microalgae *chlorella* sp. HQ for lipid production and nutrient uptake. *Water Sci. Technol.* 70, 712–719. doi: 10.2166/wst.2014.285

Zhu, Y. H., and Huang, Y. J. (2017). Use of flux balance analysis to promote lipid productivity in *Chlorella sorokiniana*. *J. Appl. Phycol.* 29, 889–902. doi: 10.1007/s10811-016-0973-6

Zhu, B. H., Zhang, R. H., Lv, N. N., Yang, G. P., Wang, Y. S., and Pan, K. H. (2018). The role of malic enzyme on promoting Total lipid and fatty acid production in Phaeodactylum tricornutum. *Front. Plant Sci.* 9:826. doi: 10.3389/fpls.2018.00826

Zou, L. G., Chen, J. W., Zheng, D. L., Balamurugan, S., Li, D. W., Yang, W. D., et al. (2018). High-efficiency promoter-driven coordinated regulation of multiple metabolic nodes elevates lipid accumulation in the model microalga *Phaeodactylum tricornutum*. *Microb. Cell Factories* 17:54. doi: 10.1186/s12934-018-0906-y

Zulu, N. N., Popko, J., Zienkiewicz, K., Tarazona, P., Herrfurth, C., and Feussner, I. (2017). Heterologous co-expression of a yeast diacylglycerol acyltransferase (ScDGA1) and a plant oleosin (AtOLEO3) as an efficient tool for enhancing triacylglycerol accumulation in the marine diatom *Phaeodactylum tricornutum*. *Biotechnol. Biofuels* 10:187. doi: 10.1186/s13068-017-0874-1

frontiers | Frontiers in Microbiology

TYPE Original Research
PUBLISHED 08 November 2022
DOI 10.3389/fmicb.2022.1042437

OPEN ACCESS

EDITED BY
Xuefeng Lu,
Qingdao Institute of Bioenergy and Bioprocess Technology (CAS), China

REVIEWED BY
Satoru Watanabe,
Tokyo University of Agriculture, Japan
Shengwei Hou,
Southern University of Science and Technology, China

*CORRESPONDENCE
Stephan Klähn
stephan.klaehn@ufz.de

†PRESENT ADDRESS
Denny Popp,
Institute of Human Genetics, University of Leipzig Medical Center, Leipzig, Germany

‡These authors have contributed equally to this work and share last authorship

SPECIALTY SECTION
This article was submitted to Microbiotechnology, a section of the journal Frontiers in Microbiology

RECEIVED 12 September 2022
ACCEPTED 11 October 2022
PUBLISHED 08 November 2022

CITATION
Bozan M, Popp D, Kallies R, da Rocha UN, Klähn S and Bühler K (2022) Whole-genome sequence of the filamentous diazotrophic cyanobacterium *Tolypothrix* sp. PCC 7712 and its comparison with non-diazotrophic *Tolypothrix* sp. PCC 7601.
Front. Microbiol. 13:1042437.
doi: 10.3389/fmicb.2022.1042437

Whole-genome sequence of the filamentous diazotrophic cyanobacterium *Tolypothrix* sp. PCC 7712 and its comparison with non-diazotrophic *Tolypothrix* sp. PCC 7601

Mahir Bozan[1], Denny Popp[2†], Rene Kallies[2], Ulisses Nunes da Rocha[2], Stephan Klähn[1*‡] and Katja Bühler[1‡]

[1]Department of Solar Materials, Helmholtz-Centre for Environmental Research (UFZ), Leipzig, Germany, [2]Department of Environmental Microbiology, Helmholtz-Centre for Environmental Research (UFZ), Leipzig, Germany

Cyanobacteria are highly promising microorganisms in forthcoming biotechnologies. Besides the systematic development of molecular tools for genetic engineering, the design of chassis strains and novel reactor concepts are in focus. The latter includes capillary biofilm reactors (CBR), which offer a high surface area-to-volume ratio and very high cell densities. In this context, *Tolypothrix* sp. PCC 7712 was found to be highly suited for this reactor system due to maximal surface coverage, extraordinarily strong biofilm attachment, and high biomass formation. Here, we provide the genome sequence of *Tolypothrix* sp. PCC 7712 to potentially allow targeted strain engineering. Surprisingly, it was almost identical to an available incomplete genome draft of *Tolypothrix* sp. PCC 7601. Thus, we completely sequenced this strain as well and compared it in detail to strain PCC 7712. Comparative genome analysis revealed 257 and 80 unique protein-coding sequences for strains PCC 7601 and PCC 7712, respectively. Clustering genomes based on average nucleotide identity (ANI) and 16S rRNA homology showed 99.98% similarity and only minor distance, respectively, between the two strains in contrast to 21 other cyanobacterial genomes. Despite these high similarities, both strains differ in the ability to fix atmospheric nitrogen and show specific sequence variations, which are discussed in the paper.

KEYWORDS

next-generation sequencing, comparative genomics, cyanobacteria, *Tolypothrix*, *Fremyella diplosiphon*

Introduction

Cyanobacteria are the only prokaryotes performing oxygenic photosynthesis; i.e., oxygen is released as a side product of light-driven water oxidation. The obtained electrons are used to drive an autotrophic metabolism based on CO_2 fixation. Furthermore, multiple species are capable of fixing dinitrogen gas (N_2; Tsygankov, 2007; Bharti et al., 2017). These features form the basis for a sustainable biotech-workhorse being independent of organic carbon and reduced nitrogen compounds, which usually add significantly to the ecological footprint of biotech processes. Although numerous proof-of-concept studies show the feasibility of using cyanobacteria as solar cell factories for producing commodity products (Angermayr et al., 2015; Betterle and Melis, 2019; Hoschek et al., 2019; Xie and Lindblad, 2022), only a few examples exist where cyanobacterial biocatalysts are applied at an economic scale. All these processes are based on biomass, which needs to be harvested to extract, e.g., pigments or lipids (Jones and Mayfield, 2012; Garlapati et al., 2019; Nagappan et al., 2020). Persisting challenges for applying cyanobacteria in production processes are low productivity and reaction stability, low cell densities due to light limitation, and insufficient light energy for efficient product conversion (Posten, 2009).

Recently, a novel reactor concept has been introduced, enabling long-term and high cell-density cultivation of cyanobacteria and potentially allowing for continuous production processes. In this regard, the unicellular model strain *Synechocystis* sp. PCC 6803 was grown as a biofilm in a capillary biofilm reactor (CBR; Hoschek et al., 2019). Biofilms are surface-attached microbial communities, supported and protected by a self-produced extracellular matrix containing mainly polysaccharides and other biopolymers like proteins, DNA, or lipids. They are widespread in nature with cyanobacteria playing a key role as primary producers in complex biofilms also termed microbial mats. In a biotechnological context, biofilms can be regarded as a robust biocatalyst, naturally immobilized to a given surface, enabling continuous bioprocessing (Halan et al., 2012).

In addition to *Synechocystis*, other cyanobacteria were screened for their utilization in CBRs (Bozan et al., 2022). In this survey, the filamentous, diazotrophic strain *Tolypothrix* sp. PCC 7712 (also known as *Gloeotrichia* sp.), first isolated from a soil sample collected in New York, United States, was identified as a top-performing organism. It was superior to all other strains investigated, e.g. in biofilm biomass formation, a low biofilm detachment rate, and high surface coverage. As processes that utilize cyanobacteria as solar cell factories aim to maximize biomass and maintain the cells in the reactor system, the above-mentioned parameters are important properties of an effective and productive catalytic biofilm. Nevertheless, to enable targeted engineering, including genetic modification of this promising cyanobacterium and to develop it further to become an established chassis strain, the genome sequence of *Tolypothrix* sp. PCC 7712 is required.

In order to establish this organism as a photo-biotech workhorse, we set out to sequence and analyze the genome of *Tolypothrix* sp. PCC 7712. Surprisingly, its genome sequence showed a high similarity to the available genome sequence of *Tolypothrix* sp. PCC 7601, also known as *Fremyella diplosiphon*, which was originally isolated from a freshwater sample at another location. Here, we present a comparative genome analysis of these two strains, which present distinct deviations in gene composition and arrangement causing substantial physiological differences between both strains. As for *Tolypothrix* sp. PCC 7601 only a permanent draft genome was available, we also provided a completed genome sequence for this strain.

Materials and methods

Cultivation and maintenance of strains

Tolypothrix sp. PCC 7712 and *Tolypothrix* sp. PCC 7601 were obtained from Pasteur Culture Collection of Cyanobacteria (PCC). Both strains were maintained on an agar-solidified BG11 medium (Rippka et al., 1979) in growth chambers (INFORS) at 25 µE m^{-2} s^{-1} illumination at 30°C. For the experiments, they were transferred to liquid media, either to standard BG11 or BG11-0 (nitrate omitted) in 250 ml flasks with 20 ml of culture volume. They were incubated at constant illumination of 25 µE m^{-2} s^{-1} at 30°C without shaking.

For determining the ability of chromatic light adaptation, bacterial pre-cultures growing in BG11-0 or BG11 media were transferred to fresh media after 3 weeks, covered with gray, red, and blue foil, and cultivated for another 3 weeks under 200 µE cm^{-2} s^{-1} light-emitting diode (LED) illumination. Whole-cell absorption spectra were analyzed using a Cary 300 UV–Vis spectrophotometer (Agilent Technologies, Santa Clara, United States).

DNA isolation and quality assessment

Genomic DNA (gDNA) was isolated from *Tolypothrix* sp. PCC 7601 and PCC 7712 cells using established extraction protocols (Wilson, 2001) with few modifications. Briefly, 1 ml cell culture grown for 2 weeks was centrifuged at 13,000 *g* for 5 min, and the pellet was resuspended in 467 µl TE buffer. After addition of 100 µl Lysozyme (10 mg/ml), the pellet was resuspended by incubating it in a thermomixer (37°C, 15 min, 500 rpm), followed by the addition of 30 µl 10% SDS. The sample was incubated at the same conditions for another 15 min. Subsequently, 10 µl proteinase K (20 mg/ml) and 6 µl RNase (10 mg/ml) were added and the final solution was incubated for 1 h at 37°C while 400 rpm shaking. Then, pre-heated 80 µl 10% cetyltrimethylammonium bromide (CTAB) in 5 M NaCl solution was added together with 100 µl of pre-heated 5 M NaCl solution and incubated for 20 min at 65°C in a water bath. An equal volume of commercially obtained phenol/chloroform/isoamyl alcohol (25:24:1) was added to the mixture

before it was centrifuged at 13,000 *g* for 10 min. The upper phase was transferred to a new tube followed by the addition of an equal volume of chloroform/isoamyl alcohol (24:1). After centrifugation (13,000 *g*, 10 min), the upper phase was transferred to a new tube, and 0.7 volume of isopropanol was added to the mixture. DNA was precipitated in isopropanol solution with centrifugation (13,000 *g*, 15 min). The pellet was washed with 70% ethanol. After removing the ethanol from the mixture by centrifugation, it was left open at 37°C for 1 h in order to remove residual ethanol from the gDNA pellet. The final pellet was resuspended with ddH_2O and stored at 4°C. The quality and quantity of gDNA were checked *via* Nanodrop One^C Spectrophotometer (Thermo Fisher Scientific, Waltham, United States) at 260 and 280 nm.

Whole-genome sequencing, assembly, and annotation

For Illumina sequencing, isolated gDNA was fragmented and a sequencing library was prepared using the NEBNext® Ultra™ II DNA Library Prep Kit for Illumina® (New England Biolabs) according to the manufacturer's instructions. Sequencing was performed on an Illumina MiSeq platform with a MiSeq Reagent Kit v3 (600-cycle; Illumina). Adapter sequences from Illumina raw sequencing data (PRJNA625426 for PCC7712 and PRJNA625641 for PCC7601) were trimmed using BBDuk of the bbmap suite v38.33.[1] In addition, whole genome sequencing was performed by PCR-free Nanopore sequencing using an R9.4.1 flow cell on a MinIon MK1B device (Oxford Nanopore Technologies) controlled by MinKnow software release 19.12.5. gDNA isolates were prepared for sequencing using an SQK-LSK 109 Ligation Sequencing Kit in combination with an EXP-NBD104 Native Barcoding Expansion Kit according to the manufacturer's instructions with the following exceptions. The incubation times for the end-repair step were increased to 30 min at room temperature and 30 min at 65°C. The time for the ligation step was increased to 60 min. Raw Nanopore sequence data were base called and demultiplexed using guppy version 3.6.0 and the provided high accuracy model. Adapter sequences were trimmed using Porechop version 0.2.4.[2] Genomes were assembled using (a) unicycler v0.4.8 in hybrid mode using Illumina and Nanopore reads (Wick et al., 2017) and (b) Flye v2.8 using Nanopore reads only (Kolmogorov et al., 2020). The resulting assemblies were polished by medaka v1.0.3 using the respective Nanopore reads[3] and by four rounds of pilon v1.22 (Walker et al., 2014) using the respective Illumina reads. Final assembly quality was checked with CheckM v1.011 using the lineage-specific workflow (Parks et al., 2015) and quast v5.0.2 (Gurevich et al., 2013). After completing assemblies, final sequences were submitted to Genbank (PRJNA625426 for PCC7712 and PRJNA625641 for PCC7601) and annotated *via* PGAP pipeline.[4] Plasmid types were identified with MOB-Recon (Robertson and Nash, 2018).

Genome comparison

Both genomes were compared using Diffseq (Aggeli et al., 2018) to get information about genome variations between the two strains. MAUVE alignment (Darling et al., 2004) was applied to align chromosomes and plasmids to rearrange their initial locations for further Diffseq analyses. The genomes and generation of the annotation list were visualized using Geneious R10.0.5 (Kearse et al., 2012). To identify unique proteins BLAST RBH implemented to the Galaxy server[5] was applied by comparing the encoding nucleotide sequences. After obtaining a "first match" list, respective sequences were extracted as a fasta file and BLAST RBH was used again against this extracted list for each genome. This process was repeated three times in total to avoid errors caused by multi-copy genes.

Pairwise comparison of different genome sets

Twenty-one other cyanobacterial species listed in Supplementary Table S3, which are either well-known in the biotechnology field or compose other *Tolypothrix* species were selected from the NCBI genome database to compare them to the newly sequenced two genomes used in this study. After accessing the selected genomes, dRep (Olm et al., 2017) MASH ANI clustering was applied in the Galaxy Server https://usegalaxy.eu/. The primary clustering ANI threshold was set to 90% and the secondary clustering ANI threshold was set to 99%.

Acetylene reduction assay

The acetylene reduction assay was applied as described (Stewart et al., 1968; Yoon and Golden, 2001) with some modifications to quantify nitrogenase activity *in vivo*. The respective cultures were adapted to nitrate-omitted media (BG11-0) for 1 week. Cells were harvested and transferred to 15 ml fresh BG11-0 medium in 20 ml GC vials; cell density was adjusted based on Chl*a* content to 1.5 µM Chl*a* which was measured according to the procedure described in a previous study (Zavřel et al., 2015). Briefly, a cell pellet obtained from 1 ml of cell culture was exposed to 1 ml of 100% methanol solution for 20 min in the dark followed by measuring the

1 https://sourceforge.net/projects/bbmap/
2 https://github.com/rrwick/Porechop
3 https://github.com/nanoporetech/medaka
4 https://www.ncbi.nlm.nih.gov/genome/annotation_prok/
5 https://usegalaxy.eu/

absorbance of the supernatant at 470, 665, and 720 nm wavelength *via* a visible range spectrophotometer (Libra S11, Biochrom, Cambridge, United Kingdom). After adjusting cell density *via* Chl*a* measurement, the 5 ml headspace was filled with acetylene (0.5%), oxygen (20.9%), and nitrogen (78.6%) gas mixture. Vials were left for incubation at 50 µE m^{-2} s^{-1} for 24 h before measuring ethylene production *via* gas chromatography (TRACE 1310; Thermo Scientific, Waltham, USA). The device was equipped with a 30 m long TracePLOT TG-BOND Q+ column with 0.32 mm inner diameter and 10 µm film thickness (Thermo Scientific, Waltham, USA). The temperature of the flame ionization detector and the oven were adjusted to 300 and 60°C, respectively. A volume of 10 µl was injected *via* a Thermo TriPlus RSH autosampler. The flow rate of the carrier gas (nitrogen) was set to 10 ml min^{-1}, with a total running time of 2 min. The calibration curve was set by injecting 0.5, 1, and 2 µl of ethylene gas.

Results

Characterization and classification of the whole genome sequence of *Tolypothrix* sp. PCC 7712

The genome sequence of *Tolypothrix* sp. PCC 7712 was obtained by using a hybrid sequencing approach based on two different techniques, namely Illumina and Nanopore. The full genome sequence of *Tolypothrix* sp. PCC 7712 consisted of one large contig covering 9 Mbp and 15 smaller contigs of 0.03–0.2 Mbp in length (Table 1). The overall GC content was 40.7%.

The largest contig (CP063785) of around 9 Mbp represents the chromosome, whereas several of the other contigs were predicted to be mobilizable and thus could represent plasmids. For the latter, we used the plasmid prediction tool MOB-Recon (Robertson and Nash, 2018). It should be noted that still several of these contigs might represent parts of the chromosome, especially when considering typically essential elements such as tRNAs. While the chromosome contig CP063785 contains 78 tRNA genes, 24 additional tRNA genes are found on contig CP063788, which possibly represents a plasmid. Nevertheless, plasmids that harbor tRNA arrays have also been reported for other bacterial species (Tran et al., 2016). Altogether, a set of 7,133 genes were identified on the largest contig and 978 genes, including 24 tRNA genes were found on the smaller contigs (Table 1).

Random sequences from the obtained PCC 7712 genome were manually analyzed using BLASTN. Remarkably, they appeared to be identical to an available genome sequence, namely that of *Tolypothrix* sp. PCC 7601 even though both strains were isolated independently and at different geographical locations. To exclude possible cross-contaminations, both strains were reordered from the PCC, gDNA was isolated and the sequencing approach was performed again for both strains (resulting in the read and assembly statistics given in Supplementary Table S1). The average nucleotide identity (ANI) of both obtained genome sequences was determined using the method FastANI (Jain et al., 2018). Indeed, both strains showed 99.98% ANI, which makes the two strains closely related but also indicates particular sequence alterations. As expected from high ANI, dRep (Olm et al., 2017) MASH ANI clustering with 21 selected cyanobacterial genomes showed only minor distance between *Tolypothrix* sp. PCC 7601 and PCC 7712, which is also reflected by a phylogenetic tree based on 16S rRNA comparison (Figure 1). The most related strains based on ANI clustering and the 16S rRNA based phylogenetic tree were *Tolypothrix tenuis* PCC 7101 followed by *Tolypothrix* sp. PCC 7910.

Despite these high similarities of *Tolypothrix* sp. PCC 7712 and PCC 7601, one strain might have genes that the other does not have, or the order of genes might differ significantly. Therefore, gene order and genome synteny were examined using MAUVE to reveal potential differences. However, also in this respect both genomes showed high similarity in their genome arrangement (Figure 2).

TABLE 1 Summary of the *Tolypothrix* sp. PCC 7712 genome as obtained from Illumina and Nanopore sequencing.

Sequence name	Size (nt)	Genes	GC %	tRNA genes	Scaffold type
CP063785	9,007,860	7,133	40.8	78	Chromosome
CP063786	214,777	194	39.8	0	Conjugative
CP063787	192,899	158	39.7	0	Mobilizable
CP063788	176,706	152	39.2	24	Mobilizable
CP063789	92,234	72	41.1	0	Conjugative
CP063790	57,284	58	42.2	0	Mobilizable
CP063791	46,852	32	41.3	0	Non-mobilizable
CP063792	42,661	42	41.6	0	Mobilizable
CP063793	42,150	34	41.8	0	Mobilizable
CP063794	39,434	34	41.5	0	Mobilizable
CP063795	39,072	36	40.5	0	Mobilizable
CP063796	37,441	37	40.6	0	Mobilizable
CP063797	36,483	35	40.7	0	Non-mobilizable
CP063798	31,341	31	41.5	0	Mobilizable
CP063799	29,597	33	42.9	0	Mobilizable
CP063800	29,106	30	41.9	0	Mobilizable
Total	10,115,897	8,111	40.7	102	

Scaffold type was determined with MOB-Recon (Robertson and Nash, 2018). This module provided information about the type of the plasmid by matching them to plasmid reference databases and about plasmid transferability, replicon family, mate-pair formation, and relaxase type.

Comparative analysis of the genomes of *Tolypothrix* sp. PCC 7712 and PCC 7601 at global scale

Tolypothrix sp. PCC 7601 is the closest relative of strain PCC 7712. So far only an incomplete genome sequence of PCC 7601 containing 74 assembly gaps was available as a permanent draft

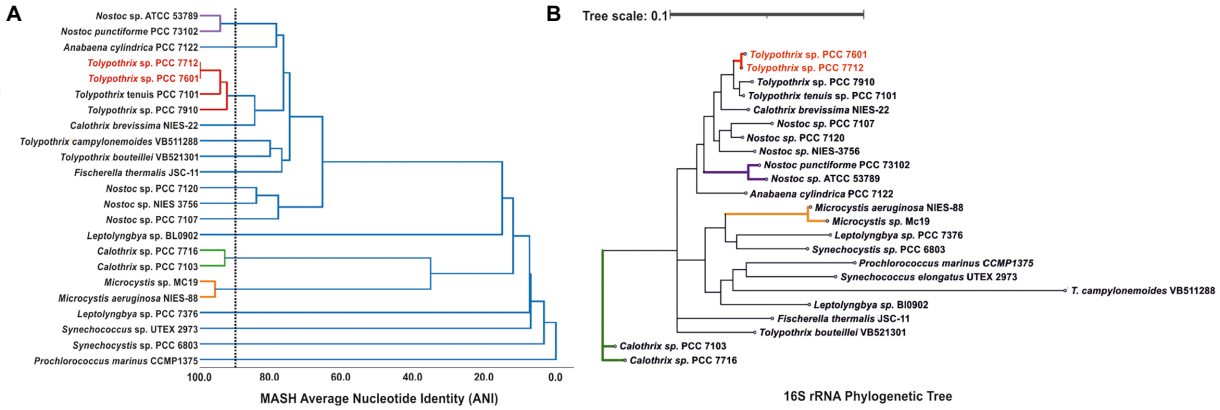

FIGURE 1
(A) The tree shows genome sequence similarity of 23 cyanobacterial strains including PCC 7601 and PCC 7712 as calculated by dRep MASH ANI clustering. The dotted line shows the primary ANI clustering threshold, which was set to 90%. **(B)** The phylogenetic tree indicates the distances between these strains based on 16S rRNA homology.

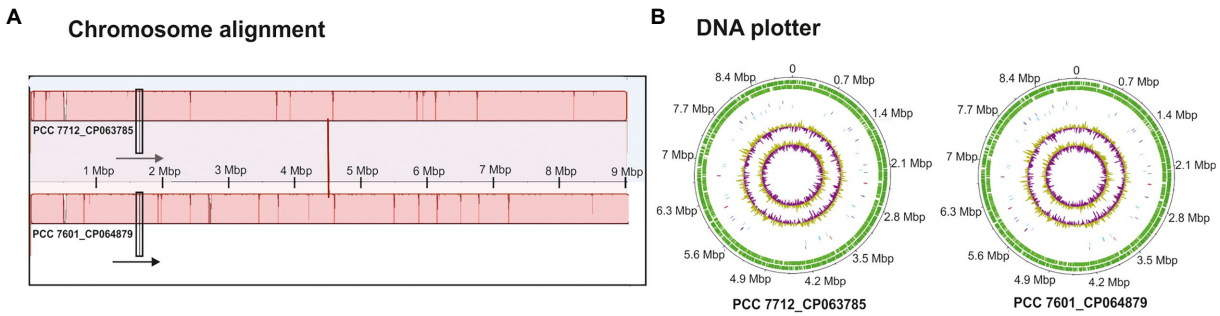

FIGURE 2
(A) Chromosome alignment together with the direction of sequences (arrows) and **(B)** DNA plotters of the two chromosomes of *Tolypothrix* sp. PCC 7712 and PCC 7601 showing GC and GC skew in the middle followed by RNAs (blue and red lines), and the outer line shows the sequence itself with its complementary strand (green).

deposited at GenBank (GCA_002368275.1). We, therefore, performed a hybrid-sequencing approach and achieved a complete circular chromosome of PCC 7601, as well as for strain PCC 7712 without any assembly gaps (Supplementary Figure S1), allowing for a detailed comparison of both strains using different bioinformatics tools.

To enable a detailed genome comparison of PCC 7712 and PCC 7601, including gene arrangement and composition, and identification of differences at gene as well as at nucleotide level, both strains' genomes were annotated using the NCBI prokaryotic genome annotation pipeline (PGAP). This resulted in the annotation given in Supplementary Table S2. The annotation revealed that strain PCC 7712 lacks one tRNA, which was identified as a tRNA-Glu. This difference appeared to originate from the insertion of an IS701 family transposase (Figure 3A). Nevertheless, the number of encoded tRNAs already indicated that there is tRNA redundancy in *Tolypothrix* genomes, which also includes multiple tRNA-Glu with the same codon usage.

Moreover, a whole-genome comparison was carried out using Diffseq (Aggeli et al., 2018), which revealed differences between both genomes in gene composition as well as at the single nucleotide level. Both strains harbor several unique genes (Table 2). For instance, the possible plasmid CP063789 of strain PCC 7712 contains an insertion that significantly alters the gene composition compared to strain PCC 7601 (Figure 3B). It contains several unique genes such as for an activating signal cointegrator 1 homology (ASCH) domain-containing protein, a helix-turn-helix domain-containing protein, an AAA family ATPase, a helicase, a DUF4186 family protein, or an ExeA family protein and several hypothetical proteins. The region of insertion showed a high query coverage to proteins also found in closely related cyanobacterial strains of the genera *Anabaena*, *Nostoc* or *Calothrix*. ExeA family protein was reported to be involved in type

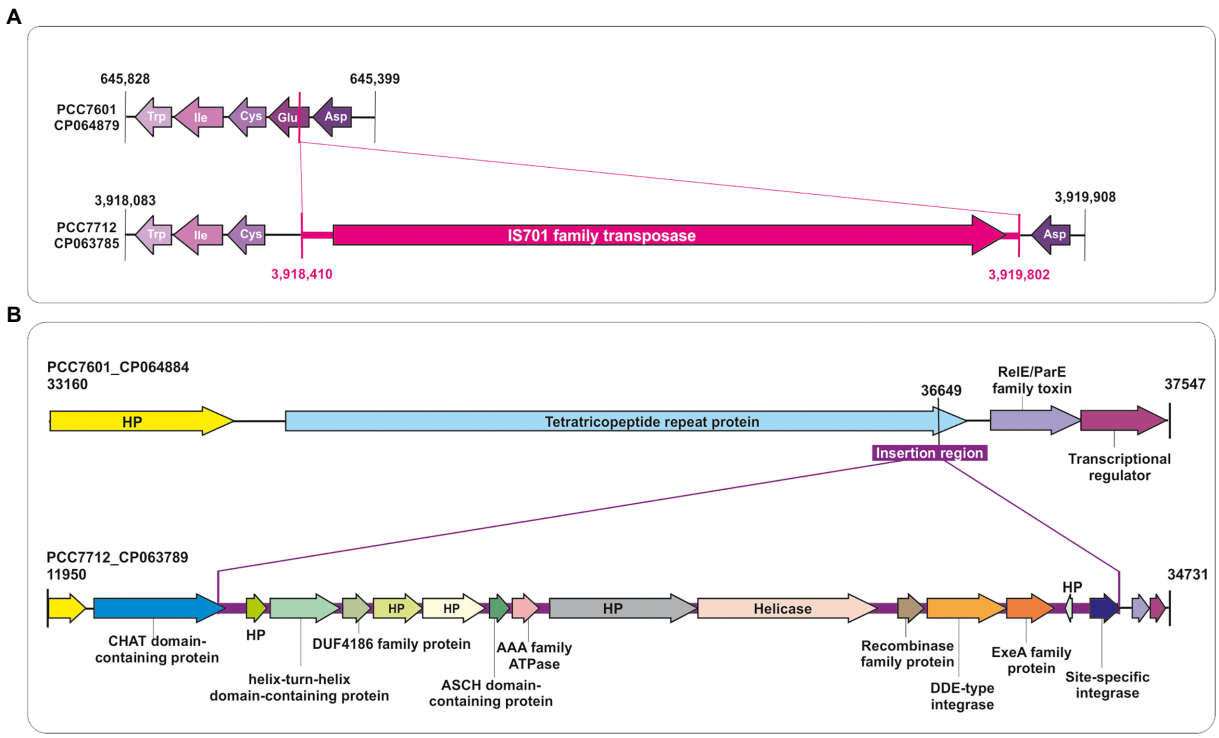

FIGURE 3
Comparative analysis of the genomes of *Tolypothrix* sp. PCC 7712 and PCC 7601 at global scale. **(A)** Insertion of a transposase belonging to the IS701 family located in the gene encoding the tRNA-Glu in the PCC 7712 chromosome. **(B)** Comparison of contigs PCC 7601_CP064884 and PCC 7712_CP063789, which differ by a large insertion sequence containing a number of protein-coding regions like five hypothetical proteins (HP), a helix-turn-helix domain-containing protein, an ASCH domain-containing protein, a DDE-type integrase, and an ExeA family protein.

II secretion systems (Vanderlinde et al., 2014), which might play a role in plasmid maintenance (Zhang et al., 2020). Further open reading frames (ORF) were uniquely identified either in the genome of strain PCC 7712 or PCC 7601 (see Table 2). Altogether, 80 unique CDS annotations were identified in strain PCC 7712 and 257 in strain PCC 7601. However, 199 CDS out of 257 (77%) in the genome of PCC 7601 and 41 out of 80 (51%) in the genome of PCC 7712 were identified as coding for hypothetical proteins (for a full list of unique ORFs see Supplementary Excel File S1).

Comparative analysis of both genomes at the single nucleotide level

Moreover, numerous single nucleotide polymorphisms (SNPs) were detected, mostly in the locations of mobile elements such as IS5, IS1634, and ISKra4 family transposases. Table 3 summarizes selected examples of the SNPs found in the genome of PCC 7712. However, most of the SNPs have been found in the non-coding regions of the genome rather than in the coding regions. These could become interesting if they affect *cis*- or *trans*-genetic elements, promotors or ribosomal binding sites. Some were also located within protein-coding regions, e.g., affecting an ATP/GTP-binding protein, HAMP domain-containing histidine kinase, and SDR family NAD(P)-dependent oxidoreductase all having a non-synonymous SNP, resulting in amino acid sequence alterations. In contrast, a gene for a tetratricopeptide repeat protein had synonymous SNP, not affecting the corresponding amino acid sequence.

Tolypothrix strains PCC 7712 and PCC 7601 are closely related but differ in nitrogen fixation ability

Microscopy analyses showed that both PCC 7712 and PCC 7601 have similar morphology (Figure 4A). Both are filamentous and the filaments are generally entangled in some regions, which explains their hairy flocs appearance in suspended batch cultures (Figures 4A,B). Furthermore, pigmentation and the ability of chromatic adaptation are fairly similar in both strains (Figures 4B,C), which is also indicated by the sequence identity in the respective proteins involved such as phytochrome superfamily photoreceptors RcaE and DpxA (Bordowitz and Montgomery, 2008; Wiltbank and Kehoe, 2016). *Tolypothrix* sp. PCC 7601 is a model organism for investigating the mechanism behind complementary chromatic acclimation (CCA; Grossman et al., 2001), while

TABLE 2 Examples of unique protein-coding regions found in the genome of PCC 7601 and PCC 7712.

Tolypothrix sp. PCC 7712			*Tolypothrix* sp. PCC 7601		
Sequence Name	**Locus_tag**	**Annotation**	**Sequence Name**	**Locus_tag**	**Annotation**
CP063785	HGR01_00210	CPXCG motif-containing cysteine-rich protein	**CP064893**	HG267_41510	3′-5′ Exonuclease
CP063785	HGR01_00215	SRPBCC family protein	**CP064879**	HG267_01625	Alpha/beta hydrolase
CP063785	HGR01_12620	Alpha/beta hydrolase	**CP064879**	HG267_09120	Carbamoylphosphate synthase large subunit
CP063785	HGR01_16155	CTB family bacteriocin	**CP064879**	HG267_25985	DDE-type integrase/transposase/ recombinase
CP063789	HGR01_38235	Helix-turn-helix domain-containing protein	**CP064879**	HG267_07765	DUF3854 domain-containing protein
CP063789	HGR01_38255	ASCH domain-containing protein	**CP064879**	HG267_32890	Formylglycine-generating enzyme family protein
CP063789	HGR01_38260	AAA family ATPase	**CP064893**	HG267_41505	HAD hydrolase-like protein
CP063789	HGR01_38270	Helicase	**CP064879**	HG267_13395	Helicase
CP063789	HGR01_38285	ExeA family protein	**CP064879**	HG267_07695	Helix-turn-helix transcriptional regulator
CP063789	HGR01_38290	Hypothetical protein	**CP064879**	HG267_07440	Histidine kinase
CP063789	HGR01_38295	Site-specific integrase	**CP064879**	HG267_00005	Hypothetical protein
CP063797	HGR01_39915	EAL domain-containing protein	**CP064879**	HG267_04935	IS5 family transposase
CP063797	HGR01_39920	Restriction endonuclease	**CP064882**	HG267_38370	IS66 family transposase
CP063797	HGR01_39925	Tyrosine-type recombinase/ integrase	**CP064879**	HG267_01590	IS701 family transposase
CP063797	HGR01_39935	GIY-YIG nuclease family protein	**CP064879**	HG267_12995	ltrA
CP063797	HGR01_39985	6-Aminohexanoate hydrolase	**CP064879**	HG267_32895	Mechanosensitive ion channel
CP063797	HGR01_40010	PT domain-containing protein	**CP064883**	HG267_39545	Relaxase/mobilization nuclease domain-containing protein
CP063797	HGR01_40025	Pentapeptide repeat-containing protein	**CP064879**	HG267_07790	Ribbon-helix–helix protein, CopG family
CP063797	HGR01_40030	Pentapeptide repeat-containing protein	**CP064893**	HG267_41535	Site-specific integrase
CP063797	HGR01_40065	AAA family ATPase	**CP064879**	HG267_25980	Tn7 transposase TnsA N-terminal domain-containing protein
CP063797	HGR01_40080	Helix-turn-helix transcriptional regulator	**CP064879**	HG267_00120	Transposase
CP063797	HGR01_40085	Integrase	**CP064879**	HG267_09560	Type II toxin-antitoxin system ParD family antitoxin
CP063798	HGR01_40105	DUF3854 domain-containing protein	**CP064879**	HG267_07845	Tyrosine-type recombinase/ integrase
CP063798	HGR01_40165	DUF3883 domain-containing protein			
CP063798	HGR01_40170	Site-specific integrase			
CP063798	HGR01_40180	DNA cytosine methyltransferase			
CP063798	HGR01_40190	ATP-binding protein			
CP063798	HGR01_40195	Type II toxin-antitoxin system VapC family toxin			
CP063798	HGR01_40210	AAA family ATPase			
CP063798	HGR01_40215	S1 RNA-binding domain-containing protein			
CP063798	HGR01_40225	TIR domain-containing protein			

The list was initially obtained by BLAST RBH and then finalized after iterative blast searches with unmatched proteins subsequently. For a full list of unique ORFs, see Supplementary Excel File S1.

TABLE 3 Examples of SNPs and insertions found in the genome of PCC 7712 compared to PCC 7601. For a complete list please refer to Supplementary Excel File S1.

Contig	Contig position (start)	Contig position (end)	Putative function	Type of variation
PCC7712_CP063785	50,753	52,296	Radical SAM protein CPXCG motif-containing cysteine-rich protein SRPBCC family protein	Insertion of 1,544 bases
PCC7712_ CP063785	231,980	233,896	ISKra4 family transposase Hypothetical protein	Insertion of 1917 bases
PCC7712_ CP063785	258,093	258,093	AAA-like domain-containing protein	Insertion of one base
PCC7712_ CP063785	286,946	287,017	Cobyrinate a,c-diamide synthase	Insertion of 72 bases
PCC7712_ CP063785	1,463,324	1,463,324	DEAD/DEAH box helicase	Insertion of one base
PCC7712_ CP063785	2,322,677	2,322,678	Helix-turn-helix domain-containing protein	Insertion of two bases
PCC7712_ CP063785	4,128,214	4,128,220	Alpha/beta hydrolase	Insertion of seven bases
PCC7712_ CP063785	6,314,548	6,314,554	Hypothetical protein	Insertion of seven bases
PCC7712_ CP063785	7,018,704	7,018,710	Iron ABC transporter permease	Insertion of seven bases
PCC7712_ CP063785	819,728	819,728	IS5 family transposase	SNP
PCC7712_ CP063785	1,218,062	1,218,062	IS1634 family transposase	SNP
PCC7712_ CP063785	1,537,608	1,537,608	ATP/GTP-binding protein	SNP
PCC7712_ CP063785	1,854,700	1,854,700	psbD	SNP
PCC7712_ CP063785	2,144,126	2,144,126	ISKra4 family transposase	SNP
PCC7712_ CP063785	2,411,161	2,411,161	23S rRNA	SNP
PCC7712_ CP063785	3,712,226	3,712,226	16S rRNA	SNP
PCC7712_ CP063785	3,996,355	3,996,355	Non-coding region of CTB family bacteriocin	SNP
PCC7712_ CP063785	3,996,451	3,996,451	CTB family bacteriocin	SNP
PCC7712_ CP063785	5,956,540	5,956,540	Tetratricopeptide repeat protein	SNP
PCC7712_ CP063785	7,391,583	7,391,583	SDR family NAD(P)-dependent oxidoreductase	SNP
PCC7712_ CP063785	8,035,613	8,035,613	HAMP domain-containing histidine kinase	SNP
PCC7712_ CP063786	23,958	23,991	DUF2127 domain-containing protein	Insertion of 34 bases
PCC7712_ CP063787	180,725	180,743	DUF1822 family protein	Insertion of 19 bases
PCC7712_ CP063788	138,690	140,082	IS701 family transposase, hypothetical protein	Insertion of 1,393 bases
PCC7712_ CP063788	62,134	62,134	Hypothetical protein	SNP
PCC7712_ CP063788	62,688	62,688	DUF1822 family protein	SNP
PCC7712_ CP063789	15,435	33,835	hypothetical protein(5x), helix-turn-helix domain-containing protein, DUF4186 family protein, ASCH domain-containing protein, AAA family ATPase, helicase, recombinase family protein, DDE-type integrase/transposase/recombinase, ExeA family protein, and site-specific integrase	Insertion of 18,401 bases
PCC7712_ CP063790	35,960	35,960	ltrA (group II intron reverse transcriptase/maturase)	Insertion of one base
PCC7712_ CP063790	45,404	45,404	TniQ family protein	Insertion of one base
PCC7712_ CP063799	27,712	27,728	ATP-binding protein	Insertion of 17 bases
PCC7712_ CP063799	21,725	21,725	ItrA (group II intron reverse transcriptase/maturase)	SNP
PCC7712_ CP063800	9,133	9,139	Protein kinase	Insertion of seven bases

this phenomenon has not been described for PCC 7712 so far, but is obviously also existing in this strain.

A major physiological difference between the two strains is the ability of PCC 7712 to fix dinitrogen gas and thus being able to grow in a medium lacking an assimilable nitrogen source like nitrate (BG11-0). Under nitrate-limiting conditions only strain PCC 7712 formed terminal heterocysts (Figure 4A; white circles), which are specialized cells harboring the nitrogenase needed for

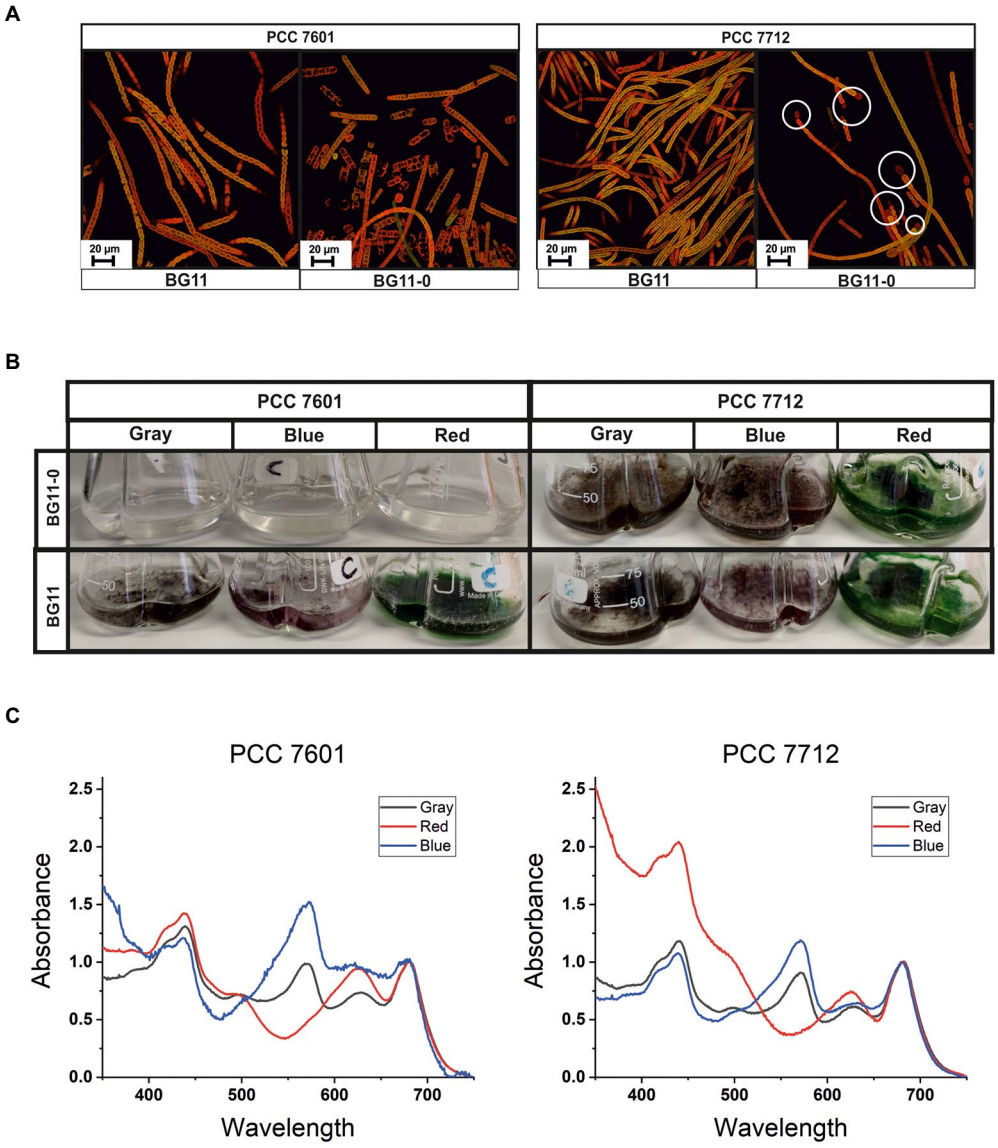

FIGURE 4
(A) CLSM images of PCC 7601 and PCC 7712 cultures growing in BG11-0 or BG11 media; white circles indicate heterocyst cells, which are less fluorescent as they lack pigmentation. **(B)** Cultures growing in different light spectra in shake flasks in either BG11-0 or BG11. **(C)** Chromatic light adaptation. Whole-cell absorption spectra of cultures covered with different colored foils—red, blue, and gray.

N_2 fixation. In contrast, strain PCC 7601 showed stressed filaments in BG11-0 lacking clear terminal heterocysts. Furthermore, no cell growth was observed which is consistent with previous reports (de Alda et al., 2004).

We performed an acetylene reduction assay to quantify the nitrogenase activity under aerobic conditions to confirm these observations. As expected, strain PCC 7601 did not show any activity, while a fairly high nitrogenase activity was detected in strain PCC 7712 converting up to 26.1 µmol H_2 mg_{Chla}^{-1} h^{-1} under here used assay conditions. Strikingly, this major difference was not observable on the genomic level based on current state-of-the-art knowledge about nitrogen fixation in *Tolypothrix*. Even though there were several unique proteins, insertions, and SNPs observed in both genomes, none of the differences were identified in the regions currently known to be related to nitrogen fixation or heterocyst formation genes (Tables 2, 3).

Discussion

Tolypothrix sp. PCC 7712 is a novel candidate strain for cyanobacterial photo-biotechnology research due to its high biofilm biomass formation, low detachment rate, and high surface coverage in a recently introduced CBR system (Bozan

et al., 2022). On the way to establish this organism as a potential chassis strain, the genome of *Tolypothrix* sp. PCC 7712 was sequenced and analyzed. We used the ANI approach, which, in contrast to the 16S rRNA-based classification, considers the total genome, and is regarded as a much more reliable tool for deciphering the degrees of bacterial relation (Jain et al., 2018). Even though strain PCC 7712 originated from a different geographical region it showed 99.9871% sequence identity with strain PCC 7601. Therefore, one could assume that both strains would have a similar phenotype despite minor differences in their genomes. This, for instance, is the case for complementary chromatic acclimation (CCA). *Via* CCA, cyanobacteria are able to change the arrangement of proteins and pigments in their phycobilisomes as well as in the accessory light-harvesting complexes connected to their photosystems when exposed to an altered light quality such as red or green light. While strain PCC 7601 has been widely used as a model strain for understanding the mechanism behind CCA (Kehoe and Gutu, 2006; Gutu and Kehoe, 2012; Montgomery, 2016), this phenomenon was so far not described for strain PCC 7712.

The most striking physiological difference was the ability of PCC7712 to fix dinitrogen gas *via* terminal heterocysts, while PCC 7601 was relying on dissolved inorganic nitrogen compounds like nitrate. This came as a surprise, as the genome of PCC 7601 harbors all essential genes for nitrogen fixation and heterocyst formation with 100% sequence identity to PCC 7712. Besides differences listed in Tables 2, 3 and Supplementary Excel File S1, there are no other differences on nucleotide level, which allow a conclusion on the reason for this observation.

Interestingly, a spontaneous revertant, *Tolypothrix* sp. PCC 7601/1 was reported previously, which is able to form heterocysts and fix dinitrogen (de Alda et al., 2004). Unfortunately, there are no sequencing data available for this strain but likely SNPs play a role in this regard. Moreover, it has been shown previously that the phycobilisome degradation protein NblA1 might play a role in the heterocyst differentiation of *Tolypothrix* sp. PCC 7601/1 but the underlying mechanism is not clearly identified yet (de Alda et al., 2004). This protein was found highly abundant under nitrogen limiting conditions in PCC 7601/1; however, it has been shown that it is also present in nitrogen replete conditions. Therefore, it was hypothesized that differential expression of the *nblA1* gene could be involved in the complex heterocyst differentiation in filamentous cyanobacteria (de Alda et al., 2004). However, when comparing both PCC 7712 and PCC 7601, we could not determine any differences neither in the *nblA1* gene nor in respective up- and downstream sequences. Moreover, there are several reports on small non-coding RNAs that are important during heterocyst differentiation in *Nostoc* sp. PCC 7120 (Brenes-Álvarez et al., 2020, 2022). Yet, a whole-genome comparison of PCC 7712 and PCC 7601 did not show any differences for small non-coding RNAs at the genome level. Nevertheless, it might be worth investigating the transcriptomes of PCC 7712 and PCC 7601 under different conditions to reveal non-coding RNAs and to unravel further details about the complex heterocyst formation process together with nitrogen fixation. Furthermore, regulatory effects based on epigenetic DNA modifications should also be considered as we identified a DNA cytosine methyltransferase (HGR01_40180) as a unique protein-coding sequence in the genome of PCC 7712 (Table 2). The presence of such a protein likely affects the DNA methylation pattern, which in turn could also affect the expression of specific genes similar to previous reports for *Synechocystis* (Gärtner et al., 2019). In this context, DNA methylation and its impact on nitrogen fixation were investigated in *Trichodesmium erythraeum* IMS101 (Walworth et al., 2017). However, in the latter cytosine methylation was not involved in the regulation of the nitrogen fixation (*nif*) genes directly. Nevertheless, the authors proposed a possible indirect effect on nitrogen fixation by influencing the expression of other nitrogen assimilatory genes. However, this is rather speculative and hence all these aspects need to be further investigated in PCC 7712 and PCC 7601. Meanwhile, there are also numerous hypothetical proteins defined uniquely in the PCC 7712 genome. Therefore, further comprehensive *in silico* analyses and physiological experiments are needed to understand the mechanism behind the difference in nitrogen fixation ability.

Conclusion

Here, we evaluated the genome of *Tolypothrix* sp. PCC 7712 which was previously introduced as a novel biofilm-forming strain with high application potential for photobiotechnological applications. Thereby, we discovered its high similarity to the genome of *Tolypothrix* sp. PCC 7601 (often referred to as *Fremyella diplosiphon*). Despite these significant similarities at the genome level, PCC 7712 was able to reduce N_2, while PCC 7601 neither shows nitrogenase activity, nor heterocyst formation. This is surprising, as the genome encodes for all necessary proteins currently known to be involved in nitrogen fixation. Thereby, these strains may become interesting models for research focusing on understanding the process of nitrogen fixation in cyanobacteria. It also shows, that a classification based on genome comparison only does not necessarily end up with close relatives also exhibiting the same physiological behavior.

Data availability statement

The datasets presented in this study are available at the Genbank database under the following accession numbers: PRJNA625426 for PCC7712 and PRJNA625641 for PCC7601.

Author contributions

MB: data curation, experiment, and writing—original draft preparation. MB, DP, RK, UR, SK, and KB: methodology and writing—review and editing. KB and SK: conceptualization, methodology, formal analysis, supervision, and writing—review and editing. All authors contributed to the article and approved the submitted version.

Acknowledgments

We acknowledge the use of the facilities of the Centre for Biocatalysis (MiKat) at the Helmholtz Centre for Environmental Research, which is supported by European Regional Development Funds (EFRE, Europe funds Saxony), and the Helmholtz Association. Cloud computing facilities used for base calling of Nanopore reads and assembly were provided by the BMBF-funded de.NBI Cloud within the German Network for Bioinformatics Infrastructure (de.NBI; 031A537B, 031A533A, 031A538A, 031A533B, 031A535A, 031A537C, 031A534A, and 031A532B).

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Supplementary material

The Supplementary material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fmicb.2022.1042437/full#supplementary-material

References

Aggeli, D., Karas, V. O., Sinnott-Armstrong, N. A., Varghese, V., Shafer, R. W., Greenleaf, W. J., et al. (2018). Diff-seq: a high throughput sequencing-based mismatch detection assay for DNA variant enrichment and discovery. *Nucleic Acids Res.* 46:e42. doi: 10.1093/nar/gky022

Angermayr, S. A., Gorchs Rovira, A., and Hellingwerf, K. J. (2015). Metabolic engineering of cyanobacteria for the synthesis of commodity products. *Trends Biotechnol.* 33, 352–361. doi: 10.1016/j.tibtech.2015.03.009

Betterle, N., and Melis, A. (2019). Photosynthetic generation of heterologous terpenoids in cyanobacteria. *Biotechnol. Bioeng.* 116, 2041–2051. doi: 10.1002/bit.26988

Bharti, A., Velmourougane, K., and Prasanna, R. (2017). Phototrophic biofilms: diversity, ecology and applications. *J. Appl. Phycol.* 29, 2729–2744. doi: 10.1007/s10811-017-1172-9

Bordowitz, J. R., and Montgomery, B. L. (2008). Photoregulation of cellular morphology during complementary chromatic adaptation requires sensor-kinase-class protein Rca E in *Fremyella diplosiphon*. *J. Bacteriol.* 190, 4069–4074. doi: 10.1128/JB.00018-08

Bozan, M., Schmid, A., and Bühler, K. (2022). Evaluation of self-sustaining cyanobacterial biofilms for technical applications. *Biofilms* 4:100073. doi: 10.1016/j.bioflm.2022.100073

Brenes-Álvarez, M., Minguet, M., Vioque, A., and Muro-Pastor, A. M. (2020). NsiR1, a small RNA with multiple copies, modulates heterocyst differentiation in the cyanobacterium Nostoc sp. PCC 7120. *Environ. Microbiol.* 22, 3325–3338. doi: 10.1111/1462-2920.15103

Brenes-Álvarez, M., Vioque, A., and Muro-Pastor, A. M. (2022). The heterocyst-specific small RNA NsiR1 regulates the commitment to differentiation in *Nostoc*. *Microbiol. Spectr.* 10:e0227421. doi: 10.1128/spectrum.02274-21

Darling, A. C. E., Mau, B., Blattner, F. R., and Perna, N. T. (2004). Mauve: multiple alignment of conserved genomic sequence with rearrangements. *Genome Res.* 14, 1394–1403. doi: 10.1101/gr.2289704

de Alda, J. A. G. O., Lichtlé, C., Thomas, J.-C., and Houmard, J. (2004). Immunolocalization of NblA, a protein involved in phycobilisome turnover, during heterocyst differentiation in cyanobacteria. *Microbiology* 150, 1377–1384. doi: 10.1099/mic.0.26992-0

Garlapati, D., Chandrasekaran, M., Devanesan, A., Mathimani, T., and Pugazhendhi, A. (2019). Role of cyanobacteria in agricultural and industrial sectors: an outlook on economically important byproducts. *Appl. Microbiol. Biotechnol.* 103, 4709–4721. doi: 10.1007/s00253-019-09811-1

Gärtner, K., Klähn, S., Watanabe, S., Mikkat, S., Scholz, I., Hess, W. R., et al. (2019). Cytosine N4-methylation via M.Ssp6803II is involved in the regulation of transcription, fine- tuning of DNA replication and DNA repair in the cyanobacterium *Synechocystis* sp. PCC 6803. *Front. Microbiol.* 10:1233. doi: 10.3389/fmicb.2019.01233

Grossman, A. R., Bhaya, D., and He, Q. (2001). Tracking the light environment by cyanobacteria and the dynamic nature of light harvesting. *J. Biol. Chem.* 276, 11449–11452. doi: 10.1074/jbc.R100003200

Gurevich, A., Saveliev, V., Vyahhi, N., and Tesler, G. (2013). QUAST: quality assessment tool for genome assemblies. *Bioinformatics* 29, 1072–1075. doi: 10.1093/bioinformatics/btt086

Gutu, A., and Kehoe, D. M. (2012). Emerging perspectives on the mechanisms, regulation, and distribution of light color acclimation in cyanobacteria. *Mol. Plant* 5, 1–13. doi: 10.1093/mp/ssr054

Halan, B., Buehler, K., and Schmid, A. (2012). Biofilms as living catalysts in continuous chemical syntheses. *Trends Biotechnol.* 30, 453–465. doi: 10.1016/j.tibtech.2012.05.003

Hoschek, A., Heuschkel, I., Schmid, A., Bühler, B., Karande, R., and Bühler, K. (2019). Mixed-species biofilms for high-cell-density application of *Synechocystis* sp. PCC 6803 in capillary reactors for continuous cyclohexane oxidation to cyclohexanol. *Bioresour. Technol.* 282, 171–178. doi: 10.1016/j.biortech.2019.02.093

Jain, C., Rodriguez-R, L. M., Phillippy, A. M., Konstantinidis, K. T., and Aluru, S. (2018). High throughput ANI analysis of 90K prokaryotic genomes reveals clear species boundaries. *Nat. Commun.* 9:5114. doi: 10.1038/s41467-018-07641-9

Jones, C. S., and Mayfield, S. P. (2012). Algae biofuels: versatility for the future of bioenergy. *Curr. Opin. Biotechnol.* 23, 346–351. doi: 10.1016/j.copbio.2011.10.013

Kearse, M., Moir, R., Wilson, A., Stones-Havas, S., Cheung, M., Sturrock, S., et al. (2012). Geneious basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data. *Bioinformatics* 28, 1647–1649. doi: 10.1093/bioinformatics/bts199

Kehoe, D. M., and Gutu, A. (2006). Responding to color: the regulation of complementary chromatic adaptation. *Annu. Rev. Plant Biol.* 57, 127–150. doi: 10.1146/annurev.arplant.57.032905.105215

Kolmogorov, M., Bickhart, D. M., Behsaz, B., Gurevich, A., Rayko, M., Shin, S. B., et al. (2020). metaFlye: scalable long-read metagenome assembly using repeat graphs. *Nat. Methods* 17, 1103–1110. doi: 10.1038/s41592-020-00971-x

Montgomery, B. L. (2016). Mechanisms and fitness implications of photomorphogenesis during chromatic acclimation in cyanobacteria. *J. Exp. Bot.* 67, 4079–4090. doi: 10.1093/jxb/erw206

Nagappan, S., Bhosale, R., Duc Nguyen, D., Pugazhendhi, A., Tsai, P.-C., Chang, S. W., et al. (2020). Nitrogen-fixing cyanobacteria as a potential resource for efficient biodiesel production. *Fuel* 279:118440. doi: 10.1016/j.fuel.2020.118440

Olm, M. R., Brown, C. T., Brooks, B., and Banfield, J. F. (2017). dRep: a tool for fast and accurate genomic comparisons that enables improved genome recovery from metagenomes through de-replication. *ISME J.* 11, 2864–2868. doi: 10.1038/ismej.2017.126

Parks, D. H., Imelfort, M., Skennerton, C. T., Hugenholtz, P., and Tyson, G. W. (2015). CheckM: assessing the quality of microbial genomes recovered from isolates, single cells, and metagenomes. *Genome Res.* 25, 1043–1055. doi: 10.1101/gr.186072.114

Posten, C. (2009). Design principles of photo-bioreactors for cultivation of microalgae. *Eng. Life Sci.* 9, 165–177. doi: 10.1002/elsc.200900003

Rippka, R., Deruelles, J., Waterbury, J. B., Herdman, M., and Stanier, R. Y. (1979). Generic assignments, strain histories and properties of pure cultures of cyanobacteria. *Microbiology* 111, 1–61. doi: 10.1099/00221287-111-1-1

Robertson, J., and Nash, J. H. E. (2018). MOB-suite: software tools for clustering, reconstruction and typing of plasmids from draft assemblies. *Microb. Genomics* 4:e000206. doi: 10.1099/mgen.0.000206

Stewart, W. D. P., Fitzgerald, G. P., and Burris, R. H. (1968). Acetylene reduction by nitrogen-fixing blue-green algae. *Arch. Mikrobiol.* 62, 336–348. doi: 10.1007/BF00425639

Tran, T. T. T., Belahbib, H., Bonnefoy, V., and Talla, E. (2016). A comprehensive tRNA genomic survey unravels the evolutionary history of tRNA arrays in prokaryotes. *Genome Biol. Evol.* 8, 282–295. doi: 10.1093/gbe/evv254

Tsygankov, A. A. (2007). Nitrogen-fixing cyanobacteria: a review. *Appl. Biochem. Microbiol.* 43, 250–259. doi: 10.1134/S0003683807030040

Vanderlinde, E. M., Zhong, S., Li, G., Martynowski, D., Grochulski, P., and Howard, S. P. (2014). Assembly of the type two secretion system in *Aeromonas hydrophila* involves direct interaction between the periplasmic domains of the assembly factor ExeB and the secretin ExeD. *PLoS One* 9:e102038. doi: 10.1371/journal.pone.0102038

Walker, B. J., Abeel, T., Shea, T., Priest, M., Abouelliel, A., Sakthikumar, S., et al. (2014). Pilon: an integrated tool for comprehensive microbial variant detection and genome assembly improvement. *PLoS One* 9:e112963. doi: 10.1371/journal.pone.0112963

Walworth, N. G., Hutchins, D. A., Dolzhenko, E., Lee, M. D., Fu, F., Smith, A. D., et al. (2017). Biogeographic conservation of the cytosine epigenome in the globally important marine, nitrogen-fixing cyanobacterium Trichodesmium. *Environ. Microbiol.* 19, 4700–4713. doi: 10.1111/1462-2920.13934

Wick, R. R., Judd, L. M., Gorrie, C. L., and Holt, K. E. (2017). Unicycler: resolving bacterial genome assemblies from short and long sequencing reads. *PLoS Comput. Biol.* 13:e1005595. doi: 10.1371/journal.pcbi.1005595

Wilson, K. (2001). Preparation of genomic DNA from bacteria. *Curr. Protoc. Mol. Biol.* 56, 2–4. doi: 10.1002/0471142727.mb0204s56

Wiltbank, L. B., and Kehoe, D. M. (2016). Two cyanobacterial photoreceptors regulate photosynthetic light harvesting by sensing teal, green, yellow, and red light. *MBio* 7, e02130–e02115. doi: 10.1128/mBio.02130-15

Xie, H., and Lindblad, P. (2022). Expressing 2-keto acid pathway enzymes significantly increases photosynthetic isobutanol production. *Microb. Cell Factories* 21:17. doi: 10.1186/s12934-022-01738-z

Yoon, H.-S., and Golden, J. W. (2001). PatS and products of nitrogen fixation control heterocyst pattern. *J. Bacteriol.* 183, 2605–2613. doi: 10.1128/JB.183.8.2605-2613.2001

Zavřel, T., Sinetova, M., and Červený, J. (2015). Measurement of chlorophyll *a* and carotenoids concentration in cyanobacteria. *Bio-Protocol* 5, 1–5. doi: 10.21769/BioProtoc.1467

Zhang, S.-P., Wang, Q., Quan, S.-W., Yu, X.-Q., Wang, Y., Guo, D.-D., et al. (2020). Type II toxin–antitoxin system in bacteria: activation, function, and mode of action. *Biophys. Rep.* 6, 68–79. doi: 10.1007/s41048-020-00109-8

frontiers | Frontiers in Microbiology

TYPE Original Research
PUBLISHED 22 December 2022
DOI 10.3389/fmicb.2022.1083387

OPEN ACCESS

EDITED BY
Xiaoming Tan,
Hubei University,
China

REVIEWED BY
Stephan Klähn,
Helmholtz Centre for Environmental Research, Helmholtz Association of German Research Centers (HZ), Germany
Qingfang He,
University of Arkansas at Little Rock, United States

*CORRESPONDENCE
Hui Wang
wanghui@qibebt.ac.cn

SPECIALTY SECTION
This article was submitted to Microbiotechnology, a section of the journal Frontiers in Microbiology

RECEIVED 29 October 2022
ACCEPTED 28 November 2022
PUBLISHED 22 December 2022

CITATION
Zhang Z, Han T, Sui J and Wang H (2022) Cryptochrome-mediated blue-light signal contributes to carotenoids biosynthesis in microalgae.
Front. Microbiol. 13:1083387.
doi: 10.3389/fmicb.2022.1083387

Cryptochrome-mediated blue-light signal contributes to carotenoids biosynthesis in microalgae

Zhongyi Zhang[1], Tianli Han[1], Jikang Sui[1] and Hui Wang[1,2]*

[1]Solar Energy Laboratory, Qingdao Institute of Bioenergy and Bioprocess Technology, Chinese Academy of Sciences (CAS), Qingdao, China, [2]Shandong Energy Research Institute, Qingdao, China

Microalgae are considered as ideal cell factories for producing natural carotenoids which display favorable biological activities. As the most important abiotic factor, light not only provides energy for photosynthetic metabolism, but also regulates numerous biological processes. Blue light is the main wavelength of light that can travel through water. Previous studies have shown that blue light triggered carotenoid accumulation in several microalgae species, but the molecular mechanism remains unclear. Cryptochromes were blue-light-absorbing photoreceptors that have been found in all studied algal genomes. In this study, several different types of cryptochrome genes were cloned from *Haematococcus pluvialis* and *Phaeodactylum tricornutum*. Among them, cryptochrome genes *HpCRY4* from *H. pluvialis* and *PtCPF1* from *P. tricornutum* were upregulated under blue light treatment, in correlation with the increase of astaxanthin and fucoxanthin contents. Besides, heterologous expression and gene knockout was performed to verify the function of *HpCRY4* and *PtCPF1* in regulating carotenoid biosynthesis in microalgae. These results indicate that carotenoid biosynthesis in microalgae promoted by blue light was mediated by cryptochromes as photoreceptors.

KEYWORDS

cryptochrome, blue light, carotenoid biosynthesis, *Phaeodactylum tricornutum*, *Haematococcus pluvialis*

Introduction

Microalgae have drawn great attention for their ability to produce a wide range of high-value-added compounds, such as proteins, carbohydrates, carotenoids and lipids, for healthy, feed additives, fuel production, and drug manufacturing (Pulz and Gross, 2004; Chacón-Lee and González-Mariño, 2010; Borowitzka, 2013). Among them, carotenoids are widely present in photoautotrophic organisms, functioning in photosynthetic light harvesting, and protecting photosynthetic apparatus against reactive oxygen species (ROS) generation (Frank and Cogdell, 1996; Hashimoto et al., 2016). Recently, carotenoids such as fucoxanthin and astaxanthin have broad commercial prospects in the fields of

nutraceutical and pharmaceutical manufacturing (Christaki et al., 2013; Galasso et al., 2017; Ávila-Román et al., 2021), making microalgae the main source for commercial carotenoids production (Dufossé et al., 2005).

Light is not only an essential source of energy, but also an important environmental signal to regulate biochemical, physiological, and behavioral processes (Spetea et al., 2014). Light quality can trigger the behavior and developmental responses of photosynthetic organisms (Mittag and Wilhelm, 2017). Blue light is the main wavelength of light that can travel through seawater and plays an important role in adapting algae to changing environmental conditions (Depauw et al., 2012). Previous studies have shown that blue light helps to increase the accumulation of carotenoids in microalgae (Katsuda et al., 2004; Lababpour et al., 2004; Suyono et al., 2015; Wang et al., 2018). Meanwhile, the expression levels of key genes involving carotenoid biosynthesis (*ZEPs*, *PSY*, *et al*) were proved to be significantly regulated by blue light strengthening (Valle et al., 2014; Yang and Wei, 2020). Although many studies demonstrated a correlation between blue light and carotenoid biosynthesis, our knowledge about the exact mechanisms is still limited.

Cryptochromes (CRYs), as blue-light-absorbing photoreceptors in plants and animals, play a critical regulatory role in plant development and the entrainment of circadian rhythms (Hoang et al., 2008; Chaves et al., 2011; Rosensweig et al., 2018). Based on evolutionary analysis, cryptochromes originated from the ancient blue-light-dependent DNA repair enzymes called photolyases, but have lost DNA repair activity and acquired a novel role in signaling (Lin, 2002; Sancar, 2003; Chaves et al., 2011). Cryptochromes were first found in *Arabidopsis thaliana* and subsequently identified ubiquitously in animals, prokaryotes, and eukaryotes. Especially, cryptochromes have been found in all studied algal genomes available so far. Algae mainly live in aquatic environments where the blue wavelengths are the dominant light components. The cryptochromes may play an important role in light signal transduction in algae. Therefore, it is reasonable to speculate that cryptochromes mediate the regulation of carotenoid biosynthesis under blue light treatment.

In this study, *Phaeodactylum tricornutum* and *Haematococcus pluvialis* cells were cultured under white and blue light, respectively. And the full-length cDNA of cryptochrome genes were cloned from *P. tricornutum* and *H. pluvialis*. The relationship between the transcription level of cryptochrome genes and important carotenoid contents in *P. tricornutum* and *H. pluvialis* was analyzed to identify the target cryptochrome genes responding to blue light. In addition, heterologous expression and gene knockout of the target cryptochrome genes were performed in *P. tricornutum* to further verify the function of regulating carotenoid accumulation. This study provides a new perspective to explain how blue light promotes carotenoid biosynthesis in microalgae.

Materials and methods

Strains and culture conditions

The *P. tricornutum* strain used in this study, UTEX 646, was obtained from the Culture Collection of Algae at the University of Texas at Austin (UTEX). The diatoms were maintained in 50 ml Erlenmeyer flasks containing 20 ml 2f liquid medium. The 2f medium contained (per liter): 300 mg $NaNO_3$, 80 mg $Na_2SiO_3{\cdot}9H_2O$, 20 mg $NaH_2PO_4{\cdot}H_2O$, 4.36 mg $Na_2EDTA{\cdot}2H_2O$, 3.15 mg $FeCl_3{\cdot}6H_2O$, 9.8 μg $CuSO_4{\cdot}5H_2O$, 22 μg $ZnSO_4{\cdot}7H_2O$, 189 μg $MnCl_2{\cdot}4H_2O$, 7 μg $Na_2MoO_4{\cdot}2H_2O$, 12 μg $CoCl_2{\cdot}6H_2O$. The medium was prepared with filtrated seawater and autoclaved at 121°C for 20 min. Cultures were kept at 23 ± 1°C, with a continuous irradiance of 30 μmol m^{-2} s^{-1}. The solid 2f medium for transformants selection was prepared by supplementing 1% (w/v) agar (Solarbio, China). After autoclaving, the medium was cooled to 60°C and mixed with 100 μg mL^{-1} zeocin before distribution into Petri dishes.

The *H. pluvialis* strain, SCCA-PK0084, was obtained from the Scandinavian Culture Center for Algae and Protozoa (SCCAP) at the University of Copenhagen, Denmark. The *H. pluvialis* culture was maintained in 50 ml flasks containing 20 ml sterilized BG11 medium at 23 ± 1°C and continuously illuminated with 30 μmol m^{-2} s^{-1} of light.

To evaluate the light-induced effects on the biosynthesis of carotenoids, the algal cells were grown to the mid-logarithm growth phase in 100 ml bubble columns (30 cm in height, 4 cm in diameter, 100 ml medium) with CO_2-enriched air (1.5%, v/v) flowing at a rate of 30 l h^{-1}. After that, the cells were placed without light for 48 h for dark treatment, and then they were treated with different lights (white and blue) of the same light intensity (5 ~ 6 μmol m^{-2} s^{-1}), respectively (Supplementary Figure S1). The microalgal cells at different time points were collected by centrifugation at 10,000 rpm for 5 min, and the pellets were stored at −80°C for subsequent pigments and transcription analysis.

Pigments extraction and analysis

Total pigments of *P. tricornutum* cells were extracted from dried samples with organic solvents and analyzed by the high-performance liquid chromatography (HPLC) method (Mulders et al., 2015; Chen et al., 2017; Yang and Wei, 2020). Briefly, 30 mg of freeze-dried *P. tricornutum* cells were ground into powder and completely mixed with 2 ml pre-cooled methanol/acetone (1:1, v/v). The supernatant was collected by centrifugation at 12,000 rpm for 15 min. Repeat addition of the organic solvents and centrifugation until the pellet was colorless. All supernatants were collected and dried after filtering through a 0.22 μm nylon membrane for further pigment analysis.

The pigments were analyzed through an Agilent 1,200 HPLC system (Waters, United States) equipped with an Agilent ZORBAX Eclipse XDB-C18 chromatographic column (5 μm particle size,

250 × 4.6 mm; Agilent, United States). The mobile phase was 85% methanol and 100% ethyl acetate with a flow rate of 0.8 ml min^{-1}. The ratio increased from 100:0 to 30:70 over 16 min, maintained at 30:70 for 9 min, and then decreased back to100:0 over 10 min. The chromatogram was recorded at 450 nm.

Total pigments of *H. pluvialis* were extracted and analyzed based on the above approach with a few adjustments. Cold methanol/chloroform (2:1, v/v) was used as the organic solvent. Chloroform and double distilled water (DDW) were added into the collected supernatants until the volume ratio of chloroform: methanol: water was 10: 10: 9. The lower fluid containing pigment and chloroform was collected by centrifugation at 8,000 rpm for 5 min and filtered through a 0.22 μm membrane. The mobile phase of HPLC was 100% methanol and 100% acetonitrile, and their ratio was 75:25. The chromatogram was recorded at 480 nm.

Cloning of cryptochrome genes from microalgae

Genome-wide analysis was performed to search the cryptochrome genes in *P. tricornutum* and *H. pluvialis* genomes based on the conserved domain of cryptochrome. And then, specific primers were designed to amplify the full-length cDNA of cryptochrome genes in *P. tricornutum* and *H. pluvialis*. All primers were listed in Supplementary Table S1.

Total RNA was extracted from algal cells using SparkZol Reagent AC0101 (Shandong Sparkjade Biotechnology Co., Ltd., China) according to the protocol. One μg of total RNA each sample was reversely transcribed into cDNA using Evo M-MLV II (Accurate Biotechnology, China) according to the manufacturer's instructions. The cDNA was diluted to 10 ng μL^{-1} in sterile distilled water (SDW) for PCR template or mRNA abundance quantification. The full-length cDNAs of cryptochrome genes were amplified using PrimeSTAR HS DNA polymerase (Takara, Japan) with specific primers. The PCR products were analyzed by 1% agarose gel electrophoresis, and the clear and bright bands with the correct length were recycled and ligated with pCE2 vector. The successful ligated product was introduced into *E. coli* for sequencing.

The microalgal cryptochrome protein sequences were translated according to the cDNA sequences by ORF finder and aligned with some typical cryptochromes using ClustalW 1.83 (Chenna et al., 2003). The typical cryptochrome sequences were available in GenBank[1] and their IDs were listed in Supplementary Table S2. The phylogenic tree was inferred from the protein alignments using MEGAX 10.1.8 (Tamura et al., 2007) with the neighbor-joining method (Poisson model; 1,000 bootstrap replicates).

Quantitative real time PCR

Quantitative real-time PCR (qRT-PCR) was performed on the Roche Light Cycler 480 (Roche, Switzerland) with PowerUp SYBR Green Master Mix (Thermo Fisher Scientific, United States) according to the manual, and the data were collected and analyzed using $2^{-\Delta\Delta Ct}$ method (Livak and Schmittgen, 2001) by the Excel software. Housekeeping genes *β-actin* and *α-tubulin* were used as the internal control in *P. tricornutum* and *H. pluvialis*, respectively. All primers used for qRT-PCR were listed in Supplementary Table S3. All assays were performed three times, and a reaction without reverse transcriptase was used as a negative control.

Vector construction

In-Fusion cloning technology was used to construct the vectors for gene heterologous expression. Firstly, the pPha-T1 (Zaslavskaia et al., 2001) plasmid was digested with *Eco*R I and *Xba* I restriction enzymes at 37°C for 30 min. The long fragments containing plasmid backbone were recycled through electrophoresis. The full-length cDNA of cryptochrome genes were amplified by PCR with specific primers (Supplementary Table S4) and the 5′ and 3′ ends of the cDNA were made to have the same sequence as the ends of the linearized pPha-T1 vector. After that, the full-length cDNA of cryptochrome genes and the linearized pPha-T1 vector were ligated using Vazyme ClonExpress Ultra One Step Cloning Kit (Vazyma, China) according to the manual.

CRISPR-Cas9 was used to knock out the cryptochrome gene. Guide RNAs (gRNAs) were designed by the online design tool crispor.tefor.net (Haeussler et al., 2016) based on the PAM sequence (5′-NGG-3′) in the exon region of the target gene. Then the target oligonucleotides were annealed at 95°C for 5 min, and slowly cooled down at room temperature to form the double-stranded inserts which were ligated into the *Bsa*I digested vector PtPuc3_diaCas9_sgRNA.[2] The PtPuc3_diaCas9_sgRNA vector contains a diatom-codon optimized Cas9 expression cassette fused with the *P. tricornutum* LHCF2 promoter and LHCF1 terminator, and a single-guide RNA (sgRNA) cassette driven by the *P. tricornutum* U6 promoter (Nymark et al., 2016). A *Ble* gene (Apt et al., 1996) controlled by the LHCF11 promoter and LHCF1 terminator was in the backbone of vector PtPuc3_diaCas9_sgRNA which confers resistance to the antibiotic zeocin.

1 https://www.ncbi.nlm.nih.gov/

2 https://www.addgene.org/109219/

Genetic transformation of *P. tricornutum*

The biolistic transformation was used to deliver DNA into *P. tricornutum* cells. The microparticle bombardment method was performed according to published procedures with minor modifications (Apt et al., 1996). Tungsten particles (0.7 μm diameter, BioRad) were coated with DNA in the presence of 2.5 M $CaCl_2$ and 0.1 M spermidine. *P. tricornutum* cells in the exponential growth phase were collected and spread on the 2f agar plates. The bombardment was performed by a PDS-1000/He Biolistic Particle Delivery system (Bio-Rad, United States) fitted with 1,350 psi rupture discs. Bombarded cells were collected and suspended in liquid 2f medium, and then placed without light 24 h for recovery; the suspension was plated onto solid 2f medium containing 100 μg mL^{-1} zeocin. The plates were placed at 23 ± 1°C under a continuous irradiance of 30 μmol m^{-2} s^{-1} light intensity for 2–3 weeks until the resistant colonies were visible and transferable.

Statistical analysis

All data were performed as the average of biological triplicate and shown as mean ± standard deviation (SD). GraphPad_Prism V8.4.0 software was used to plot histograms and line charts. The statistical analysis was performed by Student's t-test with SPSS V24.0 statistical software (IBM, United States), and significant differences are marked with lowercase letters ($p < 0.05$).

Results and discussion

Blue light promotes carotenoid biosynthesis in algae

Excess light can lead to photo-inhibition and reduce cell growth (Difusa et al., 2015). To avoid such a scenario, a low light of 5 ~ 6 μmol m^{-2} s^{-1} was used as a signal to induce carotenoid biosynthesis. The carotenoid contents of *P. tricornutum* and *H. pluvialis* were determined by HPLC after 300 min under different light treatments, including dark, white and blue.

Fucoxanthin and zeaxanthin content of *P. tricornutum* cells were significantly increased after white and blue light irradiation. Specifically, fucoxanthin content reached 8.38 and 10.65 mg g^{-1} under white and blue light, respectively (Figure 1A). It was 26.73 and 88% higher than that treated with dark (5.65 mg g^{-1}). Zeaxanthin content was 24.26 and 41.91 mg g^{-1} under white and blue light which was much higher than that under dark (4.83 mg g^{-1}; Figure 1B). Although these two carotenoids' contents were higher under both light irradiations, we can see that blue light was much more conducive to carotenoid accumulation in *P. tricornutum*. The same carotenoid content promotion occurred in *H. pluvialis*. Astaxanthin content under blue light reached 35.21 mg g^{-1} which was 84.06% higher than that under dark (Figure 1C). Meanwhile, zeaxanthin, as a former of astaxanthin, reached 60.01 and 78.42 mg g^{-1} under white and blue light respectively, much higher than that under dark (Figure 1D). The above results indicate that blue light irradiation promotes the accumulation of carotenoids, such as fucoxanthin, zeaxanthin and astaxanthin, in algae.

In consist with our study, blue light promoted astaxanthin accumulation in *H. pluvialis* had been reported in previous studies (Katsuda et al., 2004; Lababpour et al., 2004; Suyono et al., 2015). Besides the above carotenoids we analyzed and discussed in the present study, many other carotenoid accumulations are also related to blue light. For example, blue light elevated the beta-carotene biosynthesis in *Dunaliella salina* (Fu et al., 2013; Han et al., 2019); higher blue light intensities help to achieve the carotenoid and xanthophyll pigments enrichment in mustard, beet and parsley (Samuolienė et al., 2017). It seems that the accumulation of carotenoids induced by blue light irradiation is ubiquitous in higher plants and algae.

Identification of cryptochromes from *P. tricornutum and H. pluvialis*

As important photoreceptors for positive responses to blue light and near-ultraviolet light treatment, cryptochromes might play a role in carotenoid accumulation regulation. Therefore, genome-wide analysis was performed to search the cryptochrome genes in *P. tricornutum* and *H. pluvialis* genomes, and full-length cDNAs of seven cryptochrome-encoding genes were cloned (Figure 2A). Two cryptochrome genes were identified in *P. tricornutum* and named *PtCPF1* (Gene ID: 7201137) and *PtCPF2* (Gene ID: 7199524). Meanwhile, five cryptochrome genes were identified in *H. pluvialis* and named *HpCRY1*, *HpCRY2*, *HpCRY3*, *HpCRY4*, and *HpCRY5* (their sequences were in Supplementary Table S5) according to the *H. pluvialis* transcriptome data.

Based on the phylogenetic analyses, these seven cryptochromes can be divided into three types, animal-like cryptochrome, plant cryptochrome, and DASH-type (*Drosophila*, *Arabidopsis*, *Synechocystis*, *Homo*) cryptochrome (Figure 2B). HpCRY2, HpCRY3, and PtCPF1 have high sequence homology with animal-like cryptochrome, which evolved from (6–4) photolyase and always acts as major regulators of circadian rhythms (Todo, 1999; Coesel et al., 2009; Heijde et al., 2010; Zou et al., 2017). Therefore, it is speculated that these three cryptochromes may have similar modes of action, and act as blue light receptors to regulate cell physiology and metabolism. HpCRY1, HpCRY5, and PtCPF2 are found to be closely related to DASH-type cryptochrome. However, the DASH-type cryptochromes exhibit a variety of functions in different species and the specific function is still unclear. HpCRY4 is the only cryptochrome that is classified to plant cryptochrome, and the function of HpCRY4 may be more similar to the plant

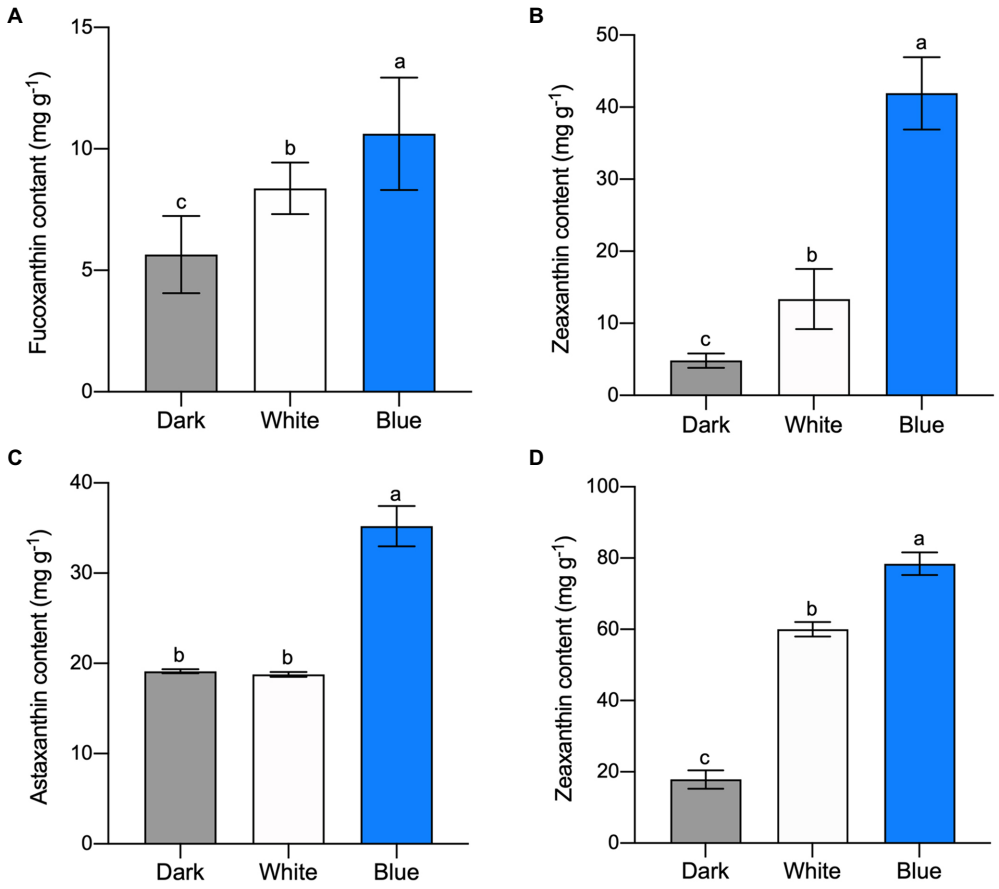

FIGURE 1
Carotenoid content in microalgae under different light irradiations. Fucoxanthin **(A)** and zeaxanthin **(B)** content in *P. tricornutum* under different light after 300min, respectively. Astaxanthin **(C)** and zeaxanthin **(D)** content in *H. pluvialis*, respectively. The vertical bars are the means ± SD of three biological replicates and lowercase letters indicate significant differences by Student's *t*-test ($p<0.05$).

cryptochrome which is involved in various growth processes such as photomorphogenesis and circadian regulation.

Compared to cryptochromes in the model plant *Arabidopsis thaliana*, there is less research on microalgal cryptochromes. Although cryptochromes from *P. tricornutum* and *H. pluvialis* were reported, respectively, in previous literatures (Petersen et al., 2021), we first dug more cryptochromes encoding genes at once and classified them into all cryptochromes types *via* bioinformatics. And it can be seen that the biological functions of algal cryptochromes are remarkably diverse among algal species. The function prediction of algal cryptochromes has certain references, but the molecular mechanisms of the relationship between cryptochrome and carotenoid biosynthesis need further analysis.

Blue light stimulates the expression of *HpCRY4* and *PtCPF1*

To validate the connection between cryptochrome and carotenoid biosynthesis in microalgae, the cryptochrome gene expression levels under white and blue lights were acquired using qRT-PCR. According to the types of cryptochromes, two cryptochrome genes from each microalga were selected for transcript abundance quantification. Among them, *HpCRY1* and *HpCRY4* were selected from five cryptochrome genes of *H. pluvialis*.

In *H. pluvialis*, the transcription level of *HpCRY4* under both light irradiations increased initially and then decreased (Figure 3A). However, compared with that under white light, the transcription level of *HpCRY4* has a more significant improvement under blue light, which was consistent with the accumulation pattern of carotenoid content. Therefore, it is speculated that *HpCRY4* probably takes part in the modulation of carotenoid-related metabolic processes and promotes astaxanthin and zeaxanthin biosynthesis in *H. pluvialis*. Meanwhile, there was no significant difference in the expression level of *HpCRY1* under different light, it seems that *HpCRY1* cannot respond to light changes at the transcriptional level.

In *P. tricornutum*, there was no statistical significance in the transcription level of *PtCPF2* under different light irradiation in 300 min (Figure 3B). The transcription level of *PtCPF1*

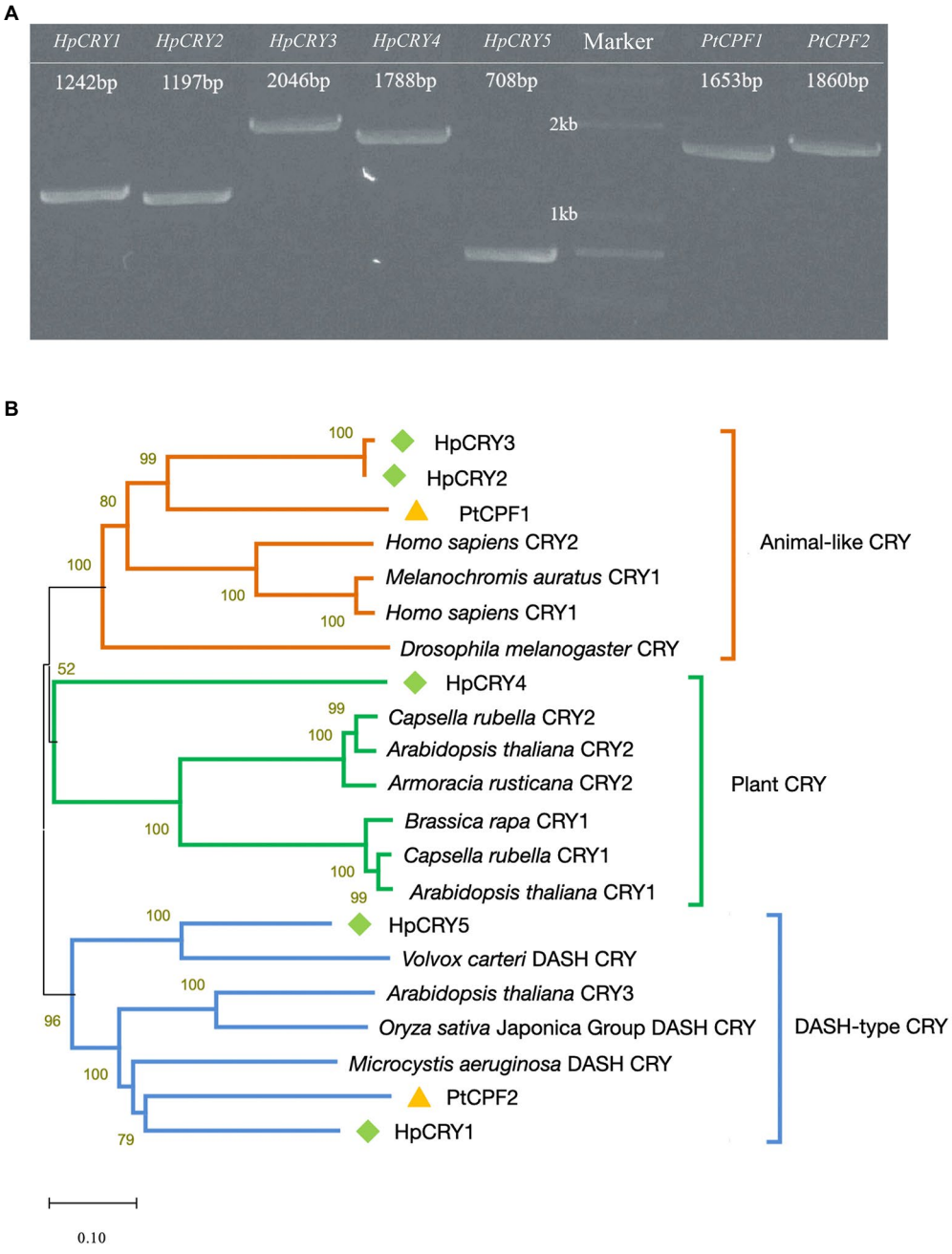

FIGURE 2
Identification of cryptochromes in *P. tricornutum* and *H. pluvialis*. **(A)** Full-length cDNA cloning of the cryptochrome genes from *P. tricornutum* and *H. pluvialis*. The numbers in panel **A** indicate the length of the sequences. **(B)** Phylogenic tree based on amino acid sequences of cryptochromes.

increased and then decreased under both light irradiations in 300 min. However, the transcription levels of *PtCPF1* under white and blue light irradiation were quite close at our sampling time point of 300 min. To further analyze the expression pattern of *PtCPF1* and *PtCPF2* and their changing trends with carotenoid contents, we reduced the induction time to 30 min and evaluated the transcription level of these two genes (Figure 3C), and fucoxanthin and zeaxanthin contents in *P. tricornutum* under different light qualities (Figure 3D) again. In 30 min, there was a minor change in the transcription level of *PtCPF2*, which indicates that it makes no response to light quality. However, the transcription level of *PtCPF1* was upregulated, and reached the highest level at 4 min, and maintained a high level in the remaining time. Interestingly, the transcription level of *PtCPF1* increased firstly and then decreased in the first 10 min and increased again in the

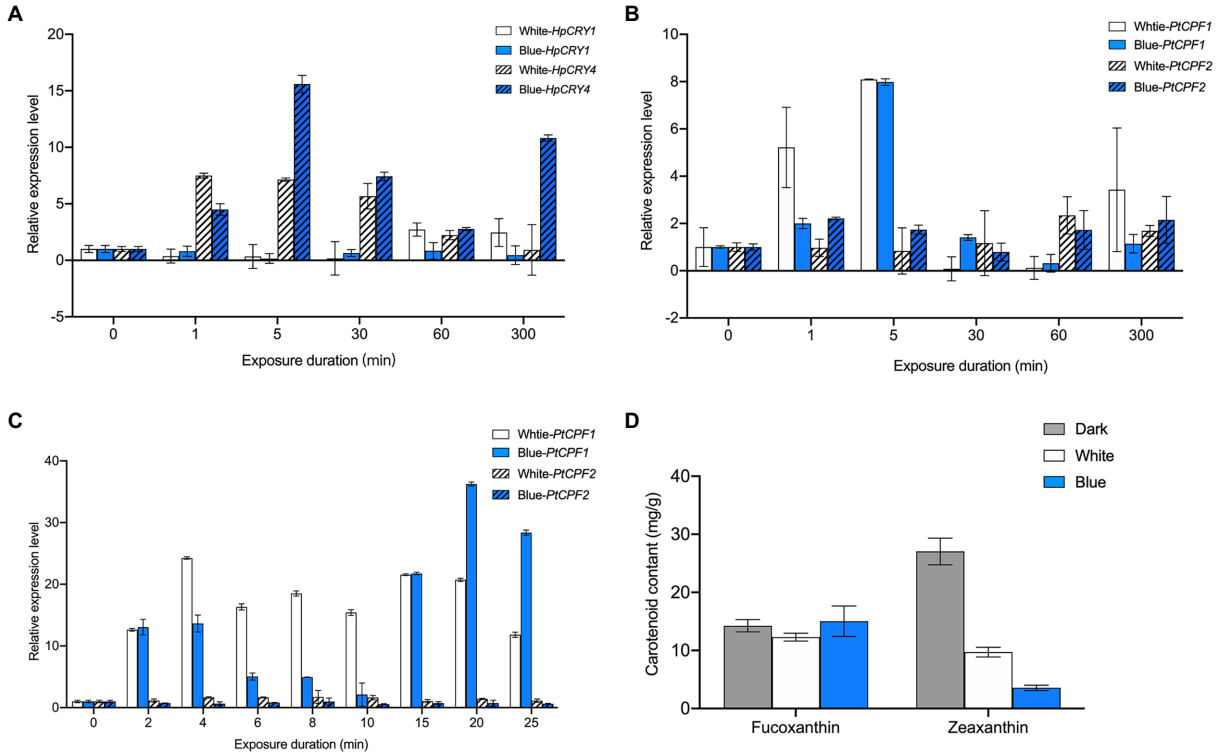

FIGURE 3
Expression analysis of cryptochrome genes under different light irradiations. **(A)** The transcription level of *HpCRY1* and *HpCRY4* in *H. pluvialis* in 300min. The expression level of *PtCPF1* and *PtCPF2* in *P. tricornutum* in 300min **(B)** and 30min **(C)**. **(D)** Fucoxanthin and zeaxanthin content in *P. tricornutum* under different light irradiation after 30min. The vertical bars are the means ± SD of three biological replicates.

10–20 min which demonstrated that the expression pattern of the *P. tricornutum PtCPF1* under blue light showed a certain periodicity. The carotenoid content in 30 min was quite different from that in 300 min. The fucoxanthin content under white and blue light in 30 min was 12.29 and 11.90 mg g^{-1}, respectively, which were not significantly different from the dark group (14.25 mg g^{-1}; $p > 0.05$). We speculated that the synthesis of fucoxanthin required a relatively complex metabolic process, and it takes time to accumulate. While the content of zeaxanthin decreased significantly after 30 min under white and blue light irradiation. Zeaxanthin would be consumed for metabolic activities in a short period when the dark-adapted *P. tricornutum* was exposed to white and blue light, and zeaxanthin consumption under blue light was higher than that under white light. The zeaxanthin accumulation may accrue in the subsequent phase when *P. tricornutum* has adapted to the light.

From the above mentioned, the plant cryptochrome gene *HpCRY4* of *H. pluvialis* and the animal-like cryptochrome gene *PtCPF1* of *P. tricornutum* can sensitively respond to the illumination variations, and their transcription levels are significantly upregulated under white and blue light irradiation. Besides, blue light has a stronger effect on *HpCRY4* and *PtCPF1* expression than white light. Although the carotenoid content is consistent with the expression patterns of *HpCRY4* and *PtCPF1* genes, whether there is a regulatory relationship between *HpCRY4* and *PtCPF1* genes and target carotenoids biosynthesis in algae requires further analysis.

Heterologous expression and knockout of some cryptochrome genes result in altered fucoxanthin content in *P. tricornutum*

The transformation and expression platforms of *H. pluvialis* have been reported in the literature (Steinbrenner and Sandmann, 2006; Kathiresan et al., 2009; Sharon-Gojman et al., 2015; Yuan et al., 2019). However, the transformation system is still immature and the expression efficiency is still low, which makes it difficult for us to characterize the function of *H. pluvialis* cryptochrome genes. In contrast, as an extensively studied model alga, *P. tricornutum* has a more mature transformation system compared with *H. pluvialis*. Considering that astaxanthin and fucoxanthin share the upstream part of β-carotene synthesis, we tried to heterologous expression of *H. pluvialis* cryptochrome genes in *P. tricornutum* to characterize the gene function.

To further explore the connection between the cryptochrome and carotenoid biosynthesis, several heterologous expression and

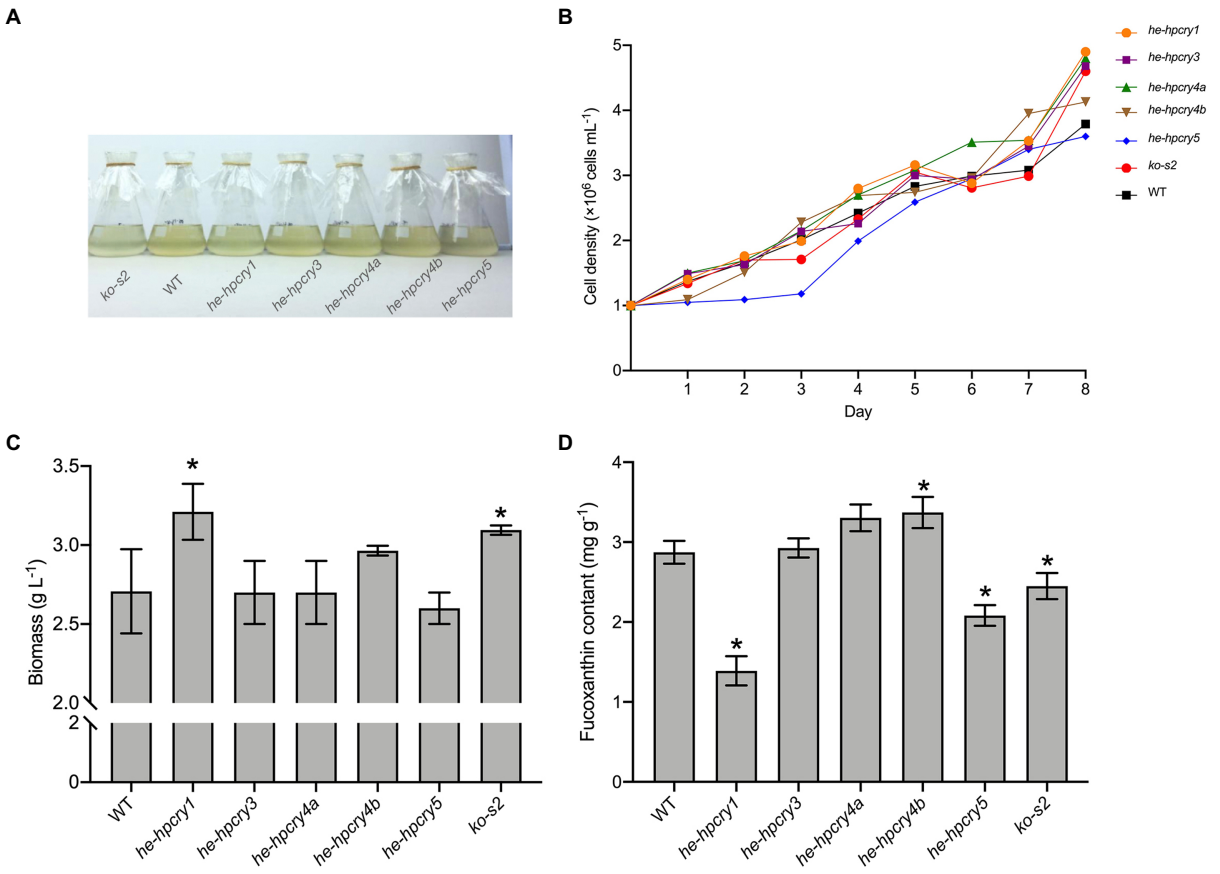

FIGURE 4
Heterologous expression and gene knockout of cryptochrome genes in *P. tricornutum*. **(A)** Cryptochrome genes' heterologous expression and knockout mutants. **(B)** The growth curve of the mutants and WT. **(C)** Total biomass of the mutants and WT. **(D)** Fucoxanthin content of the mutants and WT. The vertical bars are the means ± SD of three biological replicates. Asterisks represent significant differences from the wild type ($p<0.05$).

knockout mutants of cryptochrome genes were generated. Full-length cDNA of the *H. pluvialis* cryptochrome genes were cloned and ligated into the pPha-T1 vector. The complete recombinant plasmid was introduced into *P. tricornutum* for heterologous expression. We failed to obtain the heterologous expression mutant for *HpCRY2*, but five other *P. tricornutum* mutants were obtained including two heterologous expression mutants for *HpCRY4* (Supplementary Figure S2). These heterologous expression mutant strains were named *he-hpcry1*, *he-hpcry3*, *he-hpcry4a*, *he-hpcry4b*, and *he-hpcry5*, respectively (Figure 4A). Meanwhile, four sgRNAs were designed (Supplementary Table S6) to target the exon region of *PtCPF1* and one gene knockout mutant strain named *ko-s2* was generated by using the CRISPR/Cas9 genome editing system (Supplementary Figure S3). The sequence alignment with the wild-type *P. tricornutum* showed that thymine at position 124 was absent in *ko-s2* (Supplementary Figure S3C).

Five heterologous expression mutant strains and one knockout strain were inoculated into 250 ml flasks at an initial concentration of 1×10^6 cells mL^{-1}. The growth curve was plotted by calculating the cell concentration every 24 h. When the mutants reached the stationary phase, their biomass and fucoxanthin contents were acquired. According to the growth curve (Figure 4B), mutant *he-hpcry5* grows more slowly than the wild-type *P. tricornutum* while the other mutants have quite similar growth curves to the wild-type *P. tricornutum*. The biomass of *he-hpcry1* and *ko-s2* reached 3.21 and 3.09 g L^{-1} which were 18.61 and 14.32% higher than that of the wild type (2.71 g L^{-1}), respectively (Figure 4C; $P < 0.05$). For fucoxanthin content (Figure 4D), heterologous expression of *HpCRY1* and *HpCRY5* significantly decreased the fucoxanthin content in *P. tricornutum*. The fucoxanthin content of mutant *he-hpcry1* and *he-hpcry5* were 1.39 and 2.08 mg g^{-1} which were 51.64 and 27.53% lower than the wild type (2.87 mg g^{-1}), respectively ($p < 0.05$). Knockout of *PtCPF1* also resulted in a decrease in fucoxanthin content, and the fucoxanthin content of *ko-s2* was 2.45 mg g^{-1}, 14.63% lower than that of the wild type ($p < 0.05$). The fucoxanthin content of the two mutants for *HpCRY4* heterologous expression was 15 and 17.36% higher than that of the wild type.

Heterologous expression of *HpCRY1* in *P. tricornutum* results in high biomass and low fucoxanthin content, the function of *HpCRY1* needs further exploration. It is noteworthy that the fucoxanthin contents of mutant *he-hpcry4a* and *he-hpcry4b* both significantly increased. Previous studies have shown that cryptochrome-mediated light signaling pathways in plants and algae can affect the transcription levels of enzyme genes related to carotenoid metabolism (Coesel et al., 2009; Beel et al., 2017). Therefore, we supposed that heterologous expression of *HpCRY4* may promote the biosynthesis of upstream compounds of fucoxanthin and result in elevated fucoxanthin content in mutants *he-hpcry4a* and *he-hpcry4b*.

According to the transcription analysis, the expression level of the *P. tricornutum PtCPF1* under blue light was upregulated which consists with the increase of carotenoid content. Therefore, we hypothesized that a lower expression level of *PtCPF1* would cause a decrease in fucoxanthin accumulation. And in line with our theorized inference, the fucoxanthin content decreased in the *PtCPF1* knockout mutant *ko-s2*. However, it should be noted that knocking out *PtCPF1* does not completely block fucoxanthin biosynthesis, which revealed that carotenoid biosynthesis in microalgae is regulated by many other networks.

In summary, *HpCRY4* and *PtCPF1* are positively responsive to light quality and their transcription levels are positively correlated with the accumulation of carotenoids. The heterologous expression and gene knockout mutants of *P. tricornutum* further confirmed that they might regulate carotenoid biosynthesis in microalgae. Therefore, these two cryptochrome encoding genes can be used in transformation modification in microalgae for carotenoid accumulation in the future.

Conclusion

In this study, fucoxanthin and astaxanthin contents in *P. tricornutum* and *H. pluvialis*, respectively, were significantly promoted by blue light irradiation. Five cryptochrome genes covering three types and two cryptochrome genes covering two types were identified from *H. pluvialis* and *P. tricornutum*, respectively. Among them, the expression level of *HpCRY4* from *H. pluvialis* and *PtCPF1* from *P. tricornutum* was stimulated by blue light irradiation, which has a positive correlation with carotenoid accumulation. Furthermore, changes in the fucoxanthin contents in the heterologous expression and knockout mutants indicated that *HpCRY4* and *PtCPF1* might be the potential blue light receptors and participate in the regulation of carotenoid biosynthesis. This study reveals that the molecular mechanisms of carotenoid biosynthesis in microalgae under blue light are dependent on cryptochrome, which plays important role in blue-light sensing systems and regulates cellular processes and can be used as candidate genes for industrial algae construction.

Data availability statement

All data relevant for interpretation of this study are presented in the article and Supplementary material. Any further information is available from the corresponding author on reasonable request.

Author contributions

ZZ and TH performed the experiments and analyzed the data. ZZ wrote the original draft. JS participated in the data analysis. HW designed and organized the study and polished the manuscript. All authors contributed to the article and approved the submitted version.

Funding

This work was financially supported by the National Key Research and Development Program of China (Grant No. 2018YFA0902500) and the Shandong Taishan Scholars Program (No. tsqn202103144).

Acknowledgments

We thank the reviewers for their suggestions and comments, which helped improve the manuscript.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

Supplementary material

The Supplementary material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fmicb.2022.1083387/full#supplementary-material

References

Apt, K. E., Kroth-Pancic, P. G., and Grossman, A. R. (1996). Stable nuclear transformation of the diatom Phaeodactylum tricornutum. *Mol. Gen. Genet.* 252, 572–579. doi: 10.1007/BF02172403.

Ávila-Román, J., García-Gil, S., Rodríguez-Luna, A., Motilva, V., and Talero, E. (2021). Anti-inflammatory and anticancer effects of microalgal carotenoids. *Mar. Drugs* 19:531. doi: 10.3390/md19100531

Beel, B., Prager, K., Spexard, M., Sasso, S., Weiss, D., Müller, N., et al. (2017). A flavin binding cryptochrome photoreceptor responds to both blue and red light in *Chlamydomonas reinhardtii*. *Plant Cell* 24, 2992–3008. doi: 10.1105/tpc.112.098947

Borowitzka, M. A. (2013). High-value products from microalgae-their development and commercialization. *J. Appl. Phycol.* 25, 743–756. doi: 10.1007/s10811-013-9983-9

Chacón-Lee, T. L., and González-Mariño, G. E. (2010). Microalgae for "healthy" foods-possibilities and challenges. *Compr. Rev. Food Sci. Food Saf.* 9, 655–675. doi: 10.1111/j.1541-4337.2010.00132.x

Chaves, I., Pokorny, R., Byrdin, M., Hoang, N., Ritz, T., Brettel, K., et al. (2011). The cryptochromes: blue light photoreceptors in plants and animals. *Annu. Rev. Plant Biol.* 62, 335–364. doi: 10.1146/annurev-arplant-042110-103759

Chen, J. H., Liu, L., and Wei, D. (2017). Enhanced production of astaxanthin by *Chromochloris zofingiensis* in a microplate-based culture system under high light irradiation. *Bioresour. Technol.* 245, 518–529. doi: 10.1016/j.biortech.2017. 08.102

Chenna, R., Sugawara, H., Koike, T., Lopez, R., Gibson, T. J., Higgins, D. G., et al. (2003). Multiple sequence alignment with the Clustal series of programs. *Nucleic Acids Res.* 31, 3497–3500. doi: 10.1093/nar/gkg500

Christaki, E., Bonos, E., Giannenas, I., and Florou-Paneri, P. (2013). Functional properties of carotenoids originating from algae. *J. Sci. Food Agric.* 93, 5–11. doi: 10.1002/jsfa.5902

Coesel, S., Mangogna, M., Ishikawa, T., Heijde, M., Rogato, A., Finazzi, G., et al. (2009). Diatom PtCPF1 is a new cryptochrome/photolyase family member with DNA repair and transcription regulation activity. *EMBO Rep.* 10, 655–661. doi: 10.1038/embor.2009.59

Depauw, F. A., Rogato, A., Ribera d'Alcalá, M., and Falciatore, A. (2012). Exploring the molecular basis of responses to light in marine diatoms. *J. Exp. Bot.* 63, 1575–1591. doi: 10.1093/jxb/ers005

Difusa, A., Talukdar, J., Kalita, M. C., Mohanty, K., and Goud, V. V. (2015). Effect of light intensity and pH condition on the growth, biomass and lipid content of microalgae Scenedesmus species. *Biofuels* 6, 37–44. doi: 10.1080/17597269.2015.1045274

Dufossé, L., Galaup, P., Yaron, A., Arad, S. M., Blanc, P., Murthy, K. N. C., et al. (2005). Microorganisms and microalgae as sources of pigments for food use: a scientific oddity or an industrial reality? *Trends Food Sci. Technol.* 16, 389–406. doi: 10.1016/j.tifs.2005.02.006

Frank, H. A., and Cogdell, R. J. (1996). Carotenoids in photosynthesis. *Photochem. Photobiol.* 63, 257–264. doi: 10.1111/j.1751-1097.1996.tb03022.x

Fu, W., Guðmundsson, O., Paglia, G., Herjólfsson, G., Andrésson, O. S., Palsson, B. O., et al. (2013). Enhancement of carotenoid biosynthesis in the green microalga *Dunaliella salina* with light-emitting diodes and adaptive laboratory evolution. *Appl. Microbiol. Biotechnol.* 97, 2395–2403. doi: 10.1007/s00253-012-4502-5

Galasso, C., Corinaldesi, C., and Sansone, C. (2017). Carotenoids from marine organisms: biological functions and industrial applications. *Antioxidants* 6:96. doi: 10.3390/antiox6040096

Haeussler, M., Schönig, K., Eckert, H., Eschstruth, A., Mianné, J., Renaud, J. B., et al. (2016). Evaluation of off-target and on-target scoring algorithms and integration into the guide RNA selection tool CRISPOR. *Genome Biol.* 17:148. doi: 10.1186/s13059-016-1012-2

Han, S. I., Kim, S., Lee, C., and Choi, Y. E. (2019). Blue-red LED wavelength shifting strategy for enhancing beta-carotene production from halotolerant microalga *Dunaliella salina*. *J. Microbiol.* 57, 101–106. doi: 10.1007/s12275-019-8420-4

Hashimoto, H., Uragami, C., and Cogdell, R. J. (2016). Carotenoids and photosynthesis. *Subcell. Biochem.* 79, 111–139. doi: 10.1007/978-3-319-39126-7_4

Heijde, M., Zabulon, G., Corellou, F., Ishikawa, T., Brazard, J., Usman, A., et al. (2010). Characterization of two members of the cryptochrome/photolyase family from *Ostreococcus tauri* provides insights into the origin and evolution of cryptochromes. *Plant Cell Environ.* 33, 1614–1626. doi: 10.1111/j.1365-3040.2010.02168.x

Hoang, N., Schleicher, E., Kacprzak, S., Bouly, J. P., Picot, M., Wu, W., et al. (2008). Human and drosophila cryptochromes are light activated by flavin photoreduction in living cells. *PLoS Biol.* 6:e160. doi: 10.1371/journal.pbio.0060160

Kathiresan, S., Chandrashekar, A., Ravishankar, G. A., and Sarada, R. (2009). Agrobacterium-mediated transformation in the green alga Haematococcus pluvialis (chlorophyceae. Volvocales). *J. Appl. Phycol.* 45, 642–649. doi: 10.1111/j.1529-8817.2009.00688.x

Katsuda, T., Lababpour, A., Shimahara, K., and Katoh, S. (2004). Astaxanthin production by *Haematococcus pluvialis* under illumination with LEDs. *Enzym. Microb. Technol.* 35, 81–86. doi: 10.1016/j.enzmictec.2004.03.016

Lababpour, A., Hada, K., Shimahara, K., Katsuda, T., and Katoh, S. (2004). Effects of nutrient supply methods and illumination with blue light emitting diodes (LEDs) on astaxanthin production by *Haematococcus pluvialis*. *J. Biosci. Bioeng.* 98, 452–456. doi: 10.1016/S1389-1723(05)00311-7

Lin, C. (2002). Blue light receptors and signal transduction. *Plant Cell* 14, S207–S225. doi: 10.1105/tpc.000646

Livak, K. J., and Schmittgen, T. D. (2001). Analysis of relative gene expression data using real-time quantitative PCR and the 2[-Delta Delta C(T)] method. *Methods* 25, 402–408. doi: 10.1006/meth.2001.1262

Mittag, M., and Wilhelm, C. (2017). Light driven reactions in model algae. *J. Plant Physiol.* 217, 1–3. doi: 10.1016/j.jplph.2017.07.010

Mulders, K. J. M., Weesepoel, Y., Bodenes, P., Lamers, P. P., Vincken, J., Martens, D. K., et al. (2015). Nitrogen-depleted chlorella zofingiensis produces astaxanthin, ketolutein and their fatty acid esters: a carotenoid metabolism study. *J. Appl. Phycol.* 27, 125–140. doi: 10.1007/s10811-014-0333-3

Nymark, M., Sharma, A. K., Sparstad, T., Bones, A. M., and Winge, P. (2016). A CRISPR/Cas9 system adapted for gene editing in marine algae. *Sci. Rep.* 6:24951. doi: 10.1038/srep24951

Petersen, J., Rredhi, A., Szyttenholm, J., Oldemeyer, S., Kottke, T., and Mittag, M. (2021). The world of algae reveals a broad variety of Cryptochrome properties and functions. *Front. Plant Sci.* 12:766509. doi: 10.3389/fpls.2021.766509

Pulz, O., and Gross, W. (2004). Valuable products from biotechnology of microalgae. *Appl. Microbiol. Biotechnol.* 65, 635–648. doi: 10.1007/s00253-004-1647-x

Rosensweig, C., Reynolds, K. A., Gao, P., Laothamatas, I., Shan, Y., Ranganathan, R., et al. (2018). An evolutionary hotspot defines functional differences between CRYPTOCHROMES. *Nat. Commun.* 9:1138. doi: 10.1038/s41467-018-03503-6

Samuolienė, G., Viršilė, A., Brazaitytė, A., Jankauskienė, J., Sakalauskienė, S., Vaštakaitė, V., et al. (2017). Blue light dosage affects carotenoids and tocopherols in microgreens. *Food Chem.* 228, 50–56. doi: 10.1016/j.foodchem.2017.01.144

Sancar, A. (2003). Structure and function of DNA photolyase and cryptochrome blue-light photoreceptors. *Chem. Rev.* 103, 2203–2238. doi: 10.1021/cr0204348

Sharon-Gojman, R., Maimon, E., Leu, S., Zarka, A., and Boussiba, S. (2015). Advanced methods for genetic engineering of Haematococcus pluvialis (Chlorophyceae, Volvocales). *Algal Res.* 10, 8–15. doi: 10.1016/j.algal.2015.03.022

Spetea, C., Rintamäki, E., and Schoefs, B. (2014). Changing the light environment: chloroplast signalling and response mechanisms. *Philos. Trans. R. Soc. B* 369:20130220. doi: 10.1098/rstb.2013.0220

Steinbrenner, J., and Sandmann, G. (2006). Transformation of the green alga *Haematococcus pluvialis* with a phytoene desaturase for accelerated astaxanthin biosynthesis. *Appl. Environ. Microbiol.* 72, 7477–7484. doi: 10.1128/Aem.01461-06

Suyono, E. A., Aminin, P. L., Mu'avatun, U., Mu'avatun, U., Ramdaniyah, F., and Rohma, E. (2015). Combination of blue, red, white, and ultraviolet lights for increasing carotenoids and biomass of microalga *Haematococcus pluvialis*. *Procedia Environ. Sci.* 28, 399–405. doi: 10.1016/j.proenv.2015.07.049

Tamura, K., Dudley, J., Nei, M.,and Kumar, S. (2007). MEGA4: Molecular Evolutionary Genetics Analysis (MEGA) software version 4.0. *Mol Biol Evol.* 24, 1596–1599. doi: 10.1093/molbev/msm092

Todo, T. (1999). Functional diversity of the DNA photolyase/blue light receptor family. *Mutat. Res.* 434, 89–97. doi: 10.1016/s0921-8777(99)00013-0

Valle, K. C., Nymark, M., Aamot, I., Hancke, K., Winge, P., Andresen, K., et al. (2014). System responses to equal doses of photosynthetically usable radiation of blue, green, and red light in the marine diatom *Phaeodactylum tricornutum*. *PLoS One* 9:e0114211. doi: 10.1371/journal.pone.0114211

Wang, S., Verma, S. K., Said, I. H., Thomsen, L., Ullrich, M. S., and Kuhnert, N. (2018). Changes in the fucoxanthin production and protein profiles in *Cylindrotheca closterium* in response to blue light-emitting diode light. *Microb. Cell Factories* 17:110. doi: 10.1186/s12934-018-0957-0

Yang, R., and Wei, D. (2020). Improving fucoxanthin production in mixotrophic culture of marine diatom *Phaeodactylum tricornutum* by LED light shift and nitrogen supplementation. *Front. Bioeng. Biotechnol.* 8:820. doi: 10.3389/fbioe.2020.00820

Yuan, G., Xu, X., Zhang, W., Zhang, W., Cui, Y., Qin, S., et al. (2019). Biolistic transformation of *Haematococcus pluvialis* with constructs based on the flanking sequences of its endogenous alpha tubulin gene. *Front. Microbiol.* 10:1749. doi: 10.3389/fmicb.2019.01749

Zaslavskaia, L. A., Lippmeier, J. C., Kroth, P. G., Grpssman, A. R., and Apt, K. E. (2001). Transformation of the diatom *Phaeodactylum tricornutum* (Bacillariophyceae) with a variety of selectable marker and reporter genes. *J. Phycol.* 36, 379–386. doi: 10.1046/j.1529-8817.2000.99164.x

Zou, Y., Wenzel, S., Müller, N., Prager, K., Jung, E. M., Kothe, E., et al. (2017). An animal-like cryptochrome1 controls the Chlamydomonas sexual cycle. *Plant Physiol.* 174, 1334–1347. doi: 10.1104/pp.17.00493

frontiers | Frontiers in Microbiology

TYPE Original Research
PUBLISHED 04 January 2023
DOI 10.3389/fmicb.2022.1091502

OPEN ACCESS

EDITED BY
Xiaoming Tan,
Hubei University,
China

REVIEWED BY
Tao Sun,
Tianjin University,
China
Jinjin Diao,
Washington University in St. Louis,
United States

*CORRESPONDENCE
Boxiang Wang
✉ wangboxiang@link-spider.com
Zixi Chen
✉ chenzx@szu.edu.cn

†These authors have contributed equally to this work and share first authorship

SPECIALTY SECTION
This article was submitted to Microbiotechnology, a section of the journal Frontiers in Microbiology

RECEIVED 07 November 2022
ACCEPTED 12 December 2022
PUBLISHED 04 January 2023

CITATION
Zhang H, Liu Q, Liang Q, Wang B, Chen Z and Wang J (2023) Expression of tardigrade disordered proteins impacts the tolerance to biofuels in a model cyanobacterium *Synechocystis sp.* PCC 6803.
Front. Microbiol. 13:1091502.
doi: 10.3389/fmicb.2022.1091502

Expression of tardigrade disordered proteins impacts the tolerance to biofuels in a model cyanobacterium *Synechocystis sp.* PCC 6803

Heao Zhang[1†], Qingyang Liu[1†], Qing Liang[2], Boxiang Wang[2*], Zixi Chen[3*] and Jiangxin Wang[3]

[1]Whittle School and Studios, Shenzhen, Guangdong, China, [2]Shenzhen Link Spider Technology Co., Ltd., Shenzhen, China, [3]Shenzhen Key Laboratory of Marine Bioresource and Eco-environmental Science, Shenzhen Engineering Laboratory for Marine Algal Biotechnology, Guangdong Provincial Key Laboratory for Plant Epigenetics, College of Life Sciences and Oceanography, Shenzhen University, Shenzhen, China

Tardigrades, known colloquially as water bears or moss piglets, are diminutive animals capable of surviving many extreme environments, even been exposed to space in low Earth orbit. Recently termed tardigrade disordered proteins (TDPs) include three families as cytoplasmic-(CAHS), secreted-(SAHS), and mitochondrial-abundant heat soluble (MAHS) proteins. How these tiny animals survive these stresses has remained relatively mysterious. Cyanobacteria cast attention as a "microbial factory" to produce biofuels and high-value-added chemicals due to their ability to photosynthesis and CO_2 sequestration. We explored a lot about biofuel stress and related mechanisms in *Synechocystis sp.* PCC 6803. The previous studies show that CAHS protein heterogenous expression in bacteria, yeast, and human cells increases desiccation tolerance in these hosts. In this study, the expression of three CAHS proteins in cyanobacterium was found to affect the tolerance to biofuels, while the tolerance to Cd^{2+} and Zn^{2+} were slightly affected in several mutants. A quantitative transcriptomics approach was applied to decipher response mechanisms at the transcriptional level further.

KEYWORDS

tardigrade disordered protein, biofuel, tolerance, RNA-Seq, *Synechocystis*, cytoplasmic abundant heat soluble

1. Introduction

Tardigrades, also known as water bears and moss piglets, are diminutive animals capable of surviving many extreme conditions, like very high/low temperatures, radiation, osmotic shock, exposure to chemicals, and very low pressure close to the vacuum of outer space that is harmful to other forms of life (Hesgrove and Boothby, 2020; Arakawa and

Numata, 2021; Boothby, 2021). How these tiny animals survive these stresses has remained elusive even though they were noticed about 250 years ago. Recently termed tardigrade disordered proteins (TDPs) included three families as cytoplasmic-(CAHS), secreted-(SAHS), and mitochondrial-(MAHS) heat soluble proteins based on their subcellular locations (Hesgrove and Boothby, 2020). All members from these families are either highly expressed constitutively or significantly induced when under desiccation. Experiments also indicate members of the TDP family in allaying cellular disruption caused by different abiotic stresses (Hesgrove and Boothby, 2020). The "glass transition" hypothesis of CAHS proteins and the connection of water contents of CAHS protein were proposed in the desiccation tolerance of tardigrades (Arakawa and Numata, 2021; Boothby, 2021).

This study primarily focuses on three of the CAHS proteins, including CAHS 106094 (DP1, Swiss-Prot: P0CU52.1), CAHS 77580 (DP7, Swiss-Prot: P0CU43.1), and CAHS 86272 (DP8, Swiss-Prot: P0CU46.1) proteins. DP1 and DP7 proteins are found to form a glass-like matrix that prevents the metabolic failure of tardigrades upon desiccation. At the same time, the expression of DP8 is highly induced during desiccation (Boothby et al., 2017). DP1 and DP7 are cytoplasmic abundant heat-soluble proteins resulting from anhydrobiosis in tardigrades, but their specific mechanisms against anhydrobiosis are still unclear (Arakawa and Numata, 2021; Boothby, 2021). Researchers proposed that the tolerance of anhydrobiosis might be because of the stabilization of CAHS proteins as vitrifying small molecules like sugars but not their direct glass transition. DP8 affects survival slightly under dehydration but does not affect survival under freezing conditions (Boothby, 2021). There is also evidence that the hetero-expression of CAHS proteins increases desiccation tolerance in bacteria and yeast (Boothby et al., 2017), and improves the tolerance of human cells to hyperosmotic conditions (Tanaka et al., 2015). However, there is no data to explore whether these proteins could increase or impact abiotic stresses other than desiccation when expressed in heterologous systems, especially in microbes.

Cyanobacteria cast attention as a "microbial factory" to produce biofuels and high-value chemicals due to their ability to photosynthesis and CO_2 sequestration (Machado and Atsumi, 2012). However, to make the process economically feasible, one major hurdle to overcome is to improve the low cell tolerance to biofuels. Thus, CAHS proteins were expressed in a model cyanobacterium, *Synechocystis sp.* PCC 6803 (hereafter *Synechocystis*), to explore the possibility of improving the abiotic stress tolerance of cyanobacteria. We explored biofuel stress and related mechanisms in *Synechocystis* (Liu et al., 2012; Qiao et al., 2012; Wang et al., 2012; Tian et al., 2013; Zhu et al., 2013; Chen et al., 2014; Pei et al., 2014; Song et al., 2014; Zhu et al., 2015). In this study, three CAHS proteins were expressed in *Synechocystis* to construct several mutants. The tolerance of the mutants to several heavy metal ions and biofuels were evaluated. To further decipher responses at the transcriptional level of CAHS-expressed mutants, we applied transcriptomic analysis to comprehend the transcriptional responses to biofuels in *Synechocystis sp.* PCC 6803. This study provides a novel direction to utilize different stress proteins for improving the tolerance to high-value bioactive chemicals, including biofuels.

2. Materials and methods

2.1. *Synechocystis* culture and treatment conditions

Synechocystis sp. PCC 6803 and mutants were cultivated in BG-11 medium (pH 7.5) under a continuous light with ~50 μmol photons m^{-1} s^{-1} at 30°C. Cell density was measured using an Epoch2 Microplate Reader (BioTek, Winooski, VT, United States) at OD_{750}. Mutants harboring the pCB-SC101 plasmid with the SpecR (aminoglycoside adenylyltransferase) gene were maintained in BG-11 medium with 20 mg/ml spectinomycin.

To test the sensitivity of the mutants to environmental stress, five metal ions, and two biofuels were added to the BG-11 medium with the following additional concentrations: 5 μM Cd^{2+}, 8 μM Co^{2+}, 7 μM Zn^{2+}, 2 μM Cu^{2+}, 180 μM Mn^{2+}, 1.5% (v/v) ethanol, and 0.25% (v/v) 1-butanol. Fresh cultures of mutants were adjusted to $OD_{750} = 0.1$ and were further diluted to $OD_{750} = 0.01$, 0.001, and 0.0001. Then, 2.5 μl of the diluted cultures were carefully dropped on the BG-11 agar plates, with the metal ions or biofuels added when preparing the agar plate. After about 1–2 weeks of cultivation, the plates were taken to evaluate the mutants' sensitivity to the stress conditions.

To measure the growth curve of mutants in liquid BG-11 medium, 2 ml of the *Synechocystis* cells with the initial $OD_{750} = 0.1$ were cultivated in a 24-well microplate, with ethanol or 1-butanol added as stress, and cell density was measured daily using an Epoch2 Microplate Reader (BioTek, Winooski, VT, United States) at OD_{750}. After cultivation for 3 days, the *Synechocystis* cells were collected, frozen in liquid nitrogen, and used for transcriptomic library preparation.

2.2. Construction of the *Synechocystis* mutants

For the expression of the three disordered proteins in *Synechocystis*, the DNA sequences of these proteins were optimized according to the codon usage of *Synechocystis* (Supplementary File 1) and fully synthesized by BGI (BGI, Shenzhen, China). The shuttle vector pCB-SC101 (Liu and Pakrasi, 2018) was generated in our laboratory using Gibson assembly according to the reference (Liu and Pakrasi, 2018). Two constitutive promoters were cloned from the upstream of *ssl0452* (*nblA1*) and *sll1321* (*atpA*), and assembled to the pCB-SC101 vector using Gibson assembly, together with the coding sequences of the three disordered proteins (DP1, 7, and 8, linked to promoters as 0452 and 1321), and transformed to *E. coli* DH5α. As previously described (Wang et al., 2016; Bi et al., 2018), the

plasmids were isolated and transformed into the WT using electro-transformation. Positive clones were cultivated on BG-11 agar plates with 10 mg/ml spectinomycin for about 2 weeks and were confirmed with colony PCR analysis (mutants referred to 0452DPx and 1321DPx).

2.3. RNA isolation, quantification, and qualification

Total RNA was extracted using a Quick-RNA Miniprep Kit (Zymo Research, CA, United States). The isolated RNA of each sample was quantified by NanoDrop (Thermo Fisher Scientific Inc., CA, United States). The degradation and contamination of RNA were monitored with 1% agarose gels. RNA integrity was assessed using the Bioanalyzer 2100 system with the RNA Nano 6000 Assay Kit (Agilent Technologies, CA, United States).

2.4. Library preparation for transcriptomic sequencing

Total RNA was used as the input for the transcriptomic sequencing library preparations. mRNA was purified from total RNA with probes (Ribo-Zero rRNA Removal Kit, Illumia, CA, United States) to remove rRNA for further sequencing. Fragmentation was performed with the existence of divalent cations under elevated temperature in the First Strand Synthesis Reaction Buffer (5X). First-strand cDNA was synthesized with random hexamer primer and M-MuLV Reverse Transcriptase, then RNaseH was added to degrade the RNA. The second strand of cDNA was synthesized with DNA polymerase I, while dUTP was used to replace the dTTP in dNTP. After converting the overhangs into blunt ends *via* exonuclease/polymerase activities and the adenylation of 3′ ends of DNA fragments, the adaptors were ligated for downstream hybridization. For constructing a strand-specific library, the second strand of cDNA containing U was degraded with the USER Enzyme. The library fragments were purified with the AMPure XP system (Beckman Coulter, Beverly, United States) to select cDNA fragments with ~370–420 bp in length. After the PCR reaction with the High-Fidelity DNA polymerase, the PCR products were purified with the AMPure XP system (Beckman Coulter, Beverly, United States). Finally, the library quality was assessed on the Agilent Bioanalyzer 2100 system (Agilent Technologies, CA, United States).

2.5. Clustering and sequencing

The index-coded samples were clustered on a cBot Cluster Generation System with the TruSeq PE Cluster Kit v3-cBot-HS (Illumia, CA, United States) according to the manufacturer's instructions. The generated library preparations were sequenced on an Illumina Novaseq platform, and 150 bp paired-end reads were generated (Novogene Co., Ltd., Beijing, China).

2.6. Transcriptomic data analysis

Raw data were firstly processed through in-house Perl/bash scripts to remove reads containing adapters, reads containing N base, and low-quality reads to generate clean data. Meanwhile, the clean data's Q20, Q30, and GC content were calculated. Clean data was aligned to the reference genome (GCF_000009725.1) using Bowtie2 (v2.2.3). The reads numbers mapped to each gene were counted using HTSeq (v0.6.1). The FPKM of each gene was calculated based on the length of the gene and read counts mapped to this gene.

Before differential gene expression analysis, the read counts were normalized by edgeR R package through one scaling normalized factor for each sequenced library. The edgeR R package was used for differential expression analysis. Genes with foldchange >2 and value of $p < 0.05$ were assigned as differentially expressed genes (DEGs).

GOseq R package was used for the Gene Ontology (GO) enrichment analysis of differentially expressed genes. KOBAS software was used to test the statistical enrichment of differential expression genes in the KEGG database. The enriched GO and KEGG terms with value of $p < 0.05$ were considered significant. Terms significantly enriched in at least one comparison were included and visualized for data interpretation, even if they were not significantly enriched in other comparisons. Data visualization was achieved using in-house R scripts.

3. Results and discussion

3.1. Construction and transformation of the TDP plasmids

The pCB-SC101 vector was derived from the endogenous plasmid pCB2.4 in *Synechocystis* (Liu and Pakrasi, 2018). All mutants constructed and primers used in this work are listed in Table 1. After successfully constructing all the TDP plasmids in *E. coli*, the plasmids were transformed into the WT by electro-transformation, including the empty pCB-SC101 plasmid as a control. During the experiment, no positive clones were obtained for 0452DP8 even after several times of transformation experiments. According to the reference, the two promoters used here were all constitutive promoters with mild expression levels (Liu and Pakrasi, 2018). They were selected for expressing TDPs to avoid possible toxicity to *Synechocystis*. As $P_{ssl0452}$ is stronger than $P_{sll1321}$, it is possible that the expression of DP8 under $P_{ssl0452}$ has toxicity and caused the failure of obtaining the 0452DP8 mutants. Finally, five TDP mutants were obtained and verified with PCR with 600–800 bp PCR products (Figure 1A) and used

TABLE 1 *Synechocystis* strains and primers used in this work.

Strain names	Strain descriptions	Strain sources
WT	The wildtype of *Synechocystis sp.* PCC 6803	Our lab
PCCV	Transformed with the empty pCB-SC101 plasmid, used as the control strain	This work
0452DP1	Transformed with the pCB-SC101 plasmid harboring $P_{ssl0452}$ and disordered protein 106,094	This work
0452DP7	Transformed with the pCB-SC101 plasmid harboring $P_{ssl0452}$ and disordered protein 77,580	This work
1321DP1	Transformed with the pCB-SC101 plasmid harboring $P_{sll1321}$ and disordered protein 106,094	This work
1321DP7	Transformed with the pCB-SC101 plasmid harboring $P_{sll1321}$ and disordered protein 77,580	This work
1321DP8	Transformed with the pCB-SC101 plasmid harboring $P_{sll1321}$ and disordered protein 86,272	This work
Primer names	**Primer sequences (5' to 3')**	**Primer usage**
p0452-F	TCTTACTGTCCCTAGTGCTTGGAACGCAATGACCCAATAACTCGTACTG	Clone the promoter of ssl0452 with homology sequence to pCB-SC101 vector for Gibson assembly
p0452-R-DP1	TATTCATATTCATGGCCTCCATCTAGGTTGCCCTCCAAGGCGACTA	Clone the promoter of *ssl0452* with homology sequence to the disordered protein 106,094 for Gibson assembly
p0452-R-DP7	TGGATTCTTGCTGATAGTTACTCATCTAGGTTGCCCTCCAAGGCGACTA	Clone the promoter of *ssl0452* with homology sequence to the disordered protein 77,580 for Gibson assembly
p0452-R-DP8	CTTCTCATATTGCTGCGACATCTAGGTTGCCCTCCAAGGCGACTA	Clone the promoter of *ssl0452* with homology sequence to the disordered protein 86,272 for Gibson assembly
p1321-F	TCTTACTGTCCCTAGTGCTTGGAACGGGAGAATTGGGGGGAAGAACCAT	Clone the promoter of sll1321 with homology sequence to pCB-SC101 vector for Gibson assembly
p1321-R-DP1	TATTCATATTCATGGCCTCCATCTAGTGACTACTAGCAAGGTGAGA	Clone the promoter of *sll1321* with homology sequence to the disordered protein 106,094 for Gibson assembly
p1321-R-DP7	TGGATTCTTGCTGATAGTTACTCATCTAGTGACTACTAGCAAGGTGAGA	Clone the promoter of *sll1321* with homology sequence to the disordered protein 77,580 for Gibson assembly
p1321-R-DP8	CTTCTCATATTGCTGCGACATCTAGTGACTACTAGCAAGGTGAGA	Clone the promoter of *sll1321* with homology sequence to the disordered protein 86,272 for Gibson assembly
DP1-F	ATGGAGGCCATGAATATGAATA	Clone the disordered protein 106,094 for Gibson assembly and colony PCR to verify the disordered protein 106,094
DP1-R	TGATGCCTGGCTCTAGTATCTCGCGTTTCATCTCGCGGTTACGTT	Clone the disordered protein 106,094 with homology sequence to pCB-SC101 vector for Gibson assembly
DP1-M	CACGTGTTTTGGCTGCGTAA	Colony PCR to verify the disordered protein 106,094
DP7-F	ATGAGTAACTATCAGCAAGAATCCA	Clone the disordered protein 77,580 for Gibson assembly and colony PCR to verify the disordered protein 77,580
DP7-R	TGATGCCTGGCTCTAGTATCTCGTTATTGGTCTGGAACTTTTCACT	Clone the disordered protein 77,580 with homology sequence to pCB-SC101 vector for Gibson assembly
DP7-M	TTGCGATAGGCCTCGGTTTT	Colony PCR to verify the disordered protein 77,580
DP8-F	ATGTCGCAGCAATATGAGAAG	Clone the disordered protein 86,272 for Gibson assembly and colony PCR to verify the disordered protein 86,272
DP8-R	TGATGCCTGGCTCTAGTATCTCAAAAAGGGACTTTATCTTCTCGC	Clone the disordered protein 86,272 with homology sequence to pCB-SC101 vector for Gibson assembly

(Continued)

TABLE 1 (Continued)

Strain names	Strain descriptions	Strain sources
DP8-M	TGGTCCTTGAAGGCCTGTTC	Colony PCR to verify the disordered protein 86,272
pCB-F	GAGATACTAGAGCCAGGCATCA	Amplify the pCB-SC101 vector for Gibson assembly
pCB-R	CCAAGCACTAGGGACAGTAAGA	Amplify the pCB-SC101 vector for Gibson assembly

Then underlined are overlapping sequences that were used for the Gibson assembly.

for downstream experiments with the engineered strain harboring the empty pCB-SC101 plasmid as a control (named as PCCV).

3.2. Stress response of *Synechocystis* mutants under multiple abiotic stress conditions

An assay with BG-11 agar plates was used to effectively test the response of the five mutants under several stress conditions. For the five metal ion conditions, no significant differences were observed for Cd^{2+}, Co^{2+}, Cu^{2+}, and Mn^{2+} except for the tolerance towards Cd^{2+} of 0452DP1 was slightly improved, while all the engineered mutants were found to be slightly sensitive to Zn^{2+} compared to PCCV (Figure 1B). No significant differences were observed for the ethanol and butanol conditions for 0452DP1 and 0452DP7. However, 1321DP1 and 1321DP7 were significantly more sensitive to selected biofuels. In contrast, 1321DP8 was more tolerant to two biofuels than PCCV (Figure 1B).

Tardigrades are dramatical for surviving rough environments such as extreme high or low temperatures and desiccation. The mechanisms behind unusual tolerance to these harsh stresses remain elusive (Chavali et al., 2017; Janis et al., 2018; Hesgrove and Boothby, 2020; Yagi-Utsumi et al., 2021; Crilly et al., 2022; Malki et al., 2022; Tanaka et al., 2022). Expression of TDP proteins increases desiccation and hyperosmotic tolerance in bacteria, yeast, and human cells (Tanaka et al., 2015; Boothby et al., 2017). Experiments indicate functional roles for TDP family members in resistance against various abiotic stresses (Hesgrove and Boothby, 2020), with various aspects of TDPs for desiccation and other tolerance, suggesting that different members of TDPs may have distinct roles in resistance against different stresses, acting as diverse stress effectors (Koshland and Tapia, 2019). In this study, DP1, DP7, and DP8 showed different roles in defending biofuel stress in cyanobacterial cells, suggesting various defensive roles of different TDPs.

There are several hypotheses for TDP functions and mechanisms. TDPs were proposed to protect cells by forming higher-order assemblies such as reversible aggregates or granules under different stress stimuli (Chavali et al., 2017). Some physiochemical properties suggest that TDP folding may be determined by many physical and chemical conditions such as electrostatic interactions, charge patterning, and expanded conformations (Crilly et al., 2022), and TDPs may experience liquid–liquid phase separations when environmental stresses happen (Janis et al., 2018). In this study, All TDP expressed mutants did not show any tolerance under heavy metal stress conditions, and no significant differences were observed for biofuel conditions in 0452DP1 and 0452DP7. In contrast, 1321DP1 and 1321DP7 were significantly more sensitive to selected biofuels, and only DP8 improved tolerance to two biofuels compared with PCCV. Our data suggest that TDPs may not help defend against heavy metal stress instead of some other abiotic stresses like biofuel. We propose that different stresses cause different cellular responses and result in different TDP folding and phase separations.

Based on the sequences and domain comparison of three TDPs, they all showed a "Disordered-Coiled coil-CAHS motif 1-CAHS motif 2-Disordered" structure, with different positions of the first disordered domain between DP8 and DP1-DP7 (Table 2). At the same time, DP1 and DP7 shared very similar patterns (Table 2). This structural difference of the N-terminal may contribute to the different biofuel tolerance effects of the expression of DP8 in *Synechocystis*.

Structures, especially the C-terminal of TDPs, determined their defensive functions. Researchers found that disordered CAHS1 proteins formed stable homo-oligomers *via* the C-terminal α-helical region and became a hydrogel against the desiccation tolerance (Yagi-Utsumi et al., 2021). Similarly, CAHS-8 oligomerizes to long fibers and gels constituted of fibers in a temperature-dependent manner with the helical domain as the core of the fibrillar structure (Malki et al., 2022). Thus, it could be concluded that conserved helical C-terminal is necessary and sufficient for filament formation and gel-transition in a stress-dependent manner by CAHS proteins as the stable cell integrity under dehydration (Tanaka et al., 2022). Besides the conserved C-terminal regions in all selected TDPs in this study, our findings also indicate that N-terminal may also contribute protective functions in abiotic stresses in cyanobacteria.

Even though we did not check desiccation tolerance in the *Synechocystis* strain, some insights could still be obtained from comparison. The disaccharide trehalose was considered to play a functional role in desiccation tolerance in many species (Boothby et al., 2017). Desiccation tolerance studies in yeast, worms, and tardigrades demonstrate that both TDPs and trehalose are the major stress effectors of desiccation tolerance (Koshland and Tapia, 2019). However, trehalose is rare or not detectable in some tardigrades, suggesting different mechanisms tardigrades to survive from desiccation by producing TDPs (Boothby, 2021). Similarly, trehalose cannot be synthesized by *Synechocystis* (Mikkat et al., 1997), and salt-adapted *Synechocystis* cells mainly

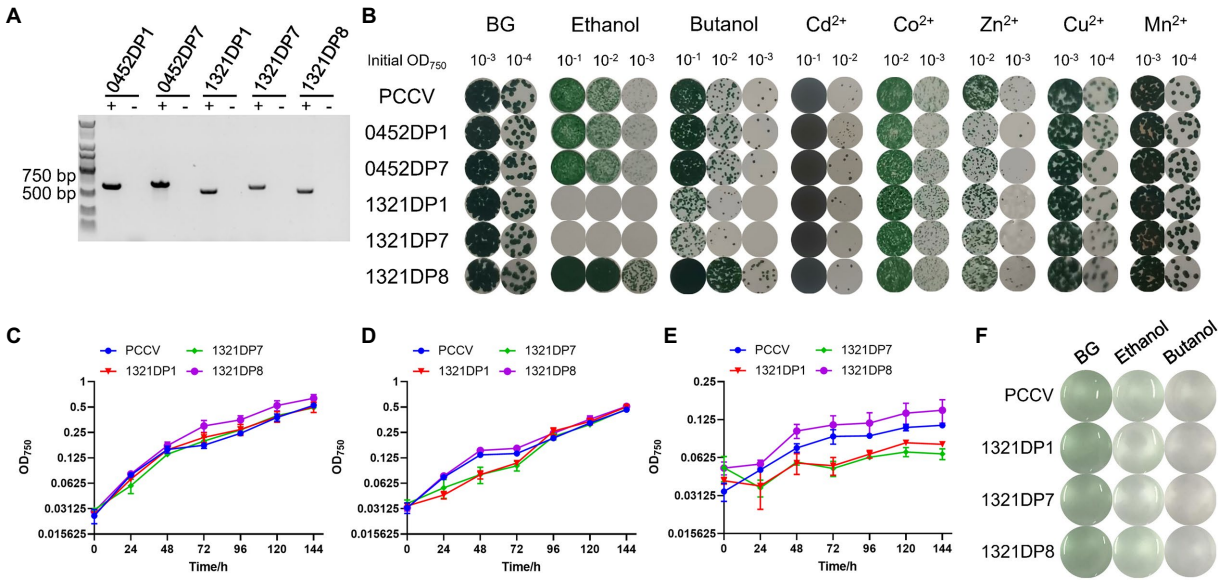

FIGURE 1
Construction and characterization of the DP mutants. **(A)** Validation of the five DP mutants by PCR. The expected amplicon size: 652bp (0452DP1), 714bp (0452DP7), 597bp (1321DP1), 659bp (1321DP7), and 585bp (1321DP8). The symbol "+" means using the corresponding DP mutants as templates, while the symbol "–" means using PCCV as negative control. **(B)** Assays of the PCCV and the five DP mutants under different stress conditions on agar plates. Cultures were adjusted to different concentrations and cultivated on agar plates with different stress conditions. **(C)** Growth curves of PCCV, 1321DP1, 1321DP7, and 1321DP8 in normal BG-11 medium. **(D)** Growth curves of PCCV, 1321DP1, 1321DP7, and 1321DP8 in normal BG-11 medium with 1.5% (v/v) ethanol. **(E)** Growth curves of PCCV, 1321DP1, 1321DP7, and 1321DP8 in normal BG-11 medium with 0.25% (v/v) 1-butanol. **(F)** PCCV, 1321DP1, 1321DP7, and 1321DP8 in normal BG-11 medium, with 1.5% (v/v) ethanol, and 0.25% (v/v) 1-butanol after 48h cultivation. The error bar represents the calculated standard deviation of the measurements of three biological replicates.

TABLE 2 Positions of family and domains section of CAHS proteins, DB1, 7, and 8, in this study.

	Disordered	Coiled coil	CAHS motif 1	CAHS motif 2	Disordered
DP1	1–28	90–140	122–140	159–177	198–227
DP7	1–38	83–191	122–140	159–177	200–224
DP8	96–125	115–193	124–142	161–179	204–237

This section provides information on sequence similarities with other proteins and the domain(s) present in a protein, based on https://www.uniprot.org/.

accumulate glucosyl glycerol (GG) and sucrose (Reed and Stewart, 1985; Hagemann et al., 1997). Thus, further elaborately designed desiccation tolerance experiments should be conducted to explore the possibility of TDPs desiccation tolerance in cyanobacteria.

3.3. Growth curve of *Synechocystis* mutants under biofuel stress

The sensitivity to biofuels of the three mutants with the $P_{sll1321}$ promoter was further confirmed with 24-well microplates. No significant differences were observed between the strains without stress (Figure 1C). Under ethanol stress, 1321DP8 showed no differences compared with PCCV, while 1321DP1 and 1321DP7 grew slower during the first 3 days, and no significant differences were found after the fourth day (Figure 1D; Supplementary Table 1). Under butanol stress, all three mutants and the PCCV control grew slowly, but 1321DP8 exhibited better tolerance than the others (Figure 1E; Supplementary Table 1). Since the biomass of cultures under butanol was very low (Figure 1F), cells cultivated under ethanol stress were collected and used for further RNA-seq.

To be noticed, slightly different phenotypes under biofuel stress were found between the results on agar plates and microplates. For example, no growth was found for 1321DP1 and 1321DP7 under ethanol agar plate, but these two mutants only grew slower during the first 3 days. The difference may be caused by different growth micro-environment of cyanobacteria. On the agar plate, not all microbes attached to the agar surface, while in the liquid medium, each cell is exposed to the microenvironments, such as heavy metals, biofuels, or salinity. The cell–cell attachments and close interactions also might be one of possible causes for the difference. Similar difference between agar plate and liquid medium were also reported in previous experiments in *Synechocystis* (Lamb and Hohmann-Marriott, 2017; Agostoni et al., 2018; Wu et al., 2020).

3.4. RNA-seq results of *Synechocystis* mutants under ethanol stress

To further reveal the differences among several *Synechocystis* mutants, four samples cultivated under ethanol stress were collected and used for transcriptomic data collection. After data trimming, over 1 GB of high-quality clean data was obtained for each sample. The clean data was then aligned to the reference genome, resulting in over 83% unique map rate for all samples (Table 3). After quantification of the expression for all the genes, the Pearson correlation coefficients between samples were calculated using the FPKM values, and 1321DP8 was found to have the lowest coefficients with other samples (Figure 2A). The quantification results confirmed the expression of TDPs in the mutants (Supplementary Table 2). The readcount values of TDPs in 1321DP1 and 1321DP8 were higher than the readcount values of *sll1321*. In 1321DP7, although only 137 reads were mapped to DP7, the FPKM value of DP7 was near 50, suggesting the expression of DP7 in 1321DP7. Statistics of DEGs also showed when compared with PCCV, 1321DP8 has more DEGs than the other two DP mutants (Figure 2B), and over 80% of the DEGs in 1321DP8 were uniquely detected (Figure 2C). In addition, the number of down-regulated DEGs was more than the number of up-regulated DEGs in all three DP mutants (Figure 2B).

The detected DEGs in the three DP mutants were used for GO and KEGG enrichment analysis, respectively, and the results were integrated and visualized for further analysis (Figure 2D). The result showed 1321DP1 down-regulated pathways involved ion and anion transport, protein modification, primary metabolic and oxidation-reduction processes, cofactor biosynthetic process, and response to stimulus in GO enrichment analysis. 1321DP7 showed similar responses with 1321DP1: with almost down-regulated processes and pathways, such as ion and anion transport, primary metabolic and oxidation–reduction processes, and cofactor biosynthetic process in GO enrichment analysis.

Synechocystis cells tend to up-regulate specific transporters and efflux pumps to eliminate the toxicity when suffering biofuel stresses. The up-regulation can be mediated by Two-component signal transduction systems (TCSs; Fu et al., 2016). As expected, sulfur metabolism and transporters pathways were KEGG enriched in 1321DP1 and 1321DP7 mutant cells, suggesting the DP1 and DP7 also trigger a similar ethanol response in cyanobacteria.

The previous RNA-seq and proteomic results showed that ethanol exposure induced genes involved in many stress responses, like heat shock proteins, transporters, as well as cell envelope modifiers, as a complicated cellular response (Wang et al., 2012), consistent with our observation in the cells upon ethanol exposure. Similarly, GO and KEGG analyses showed that many transporting, membrane-bound, oxidative stress, and sulfur relay system and photosynthesis-related proteins were induced against another biofuel hexane (Liu et al., 2012). Compared with DP1 and DP7, the expression of DP8 in cyanobacteria involved more processes and pathways. In GO enrichment analysis, more processes were up-regulated in 1321DP8 mutant cells, including signal transduction, response to stimulus, porphyrin-containing compound metabolic, phospholipid metabolic, and proteolysis. Based on KEGG enrichment, porphyrin metabolism and TCSs were significant in 1321DP8 mutant cells. Previous OMICS analyses revealed some TCS genes induced by biofuel stresses in *Synechocystis* (Fu et al., 2016), and knock-out of a TCS *slr1037* increases *Synechocystis* sensitivity to butanol (Fu et al., 2016). Compared with DP1 and DP7, DP8 might involve more in TCS for more effective tolerance against ethanol stress.

TABLE 3 Statistics of RNA-seq data.

	PCCV	1321DP1	1321DP7	1321DP8
Raw reads	7,719,810	7,941,990	8,115,158	7,615,144
Clean reads	7,632,440	7,881,858	8,052,880	7,558,308
Raw bases	1.16G	1.2G	1.22G	1.15G
Clean bases	1.15G	1.19G	1.21G	1.14G
Error rate	0.03	0.03	0.03	0.03
Q20	97.65	97.75	97.74	97.82
Q30	93.24	93.35	93.19	93.39
GC content	49.97	50.37	50.23	50.3
rRNA	0.47	0.44	0.38	0.39
Total mapped	7,366,806 (96.52%)	7,838,479 (99.45%)	7,941,202 (98.61%)	7,509,040 (99.35%)
Multiple mapped	958,925 (12.56%)	1,258,513 (15.97%)	1,182,923 (14.69%)	1,125,455 (14.89%)
Uniquely mapped	6,407,881 (83.96%)	6,579,966 (83.48%)	6,758,279 (83.92%)	6,383,585 (84.46%)
Read-1 mapped	3,206,528 (42.01%)	3,292,885 (41.78%)	3,381,947 (42%)	3,194,280 (42.26%)
Read-2 mapped	3,201,353 (41.94%)	3,287,081 (41.7%)	3,376,332 (41.93%)	3,189,305 (42.2%)
Reads map to '+'	3,204,036 (41.98%)	3,290,121 (41.74%)	3,379,175 (41.96%)	3,192,180 (42.23%)
Reads map to '–'	3,203,845 (41.98%)	3,289,845 (41.74%)	3,379,104 (41.96%)	3,191,405 (42.22%)
Reads mapped in proper pairs	6,343,264 (83.11%)	6,515,694 (82.67%)	6,698,486 (83.18%)	6,323,232 (83.66%)

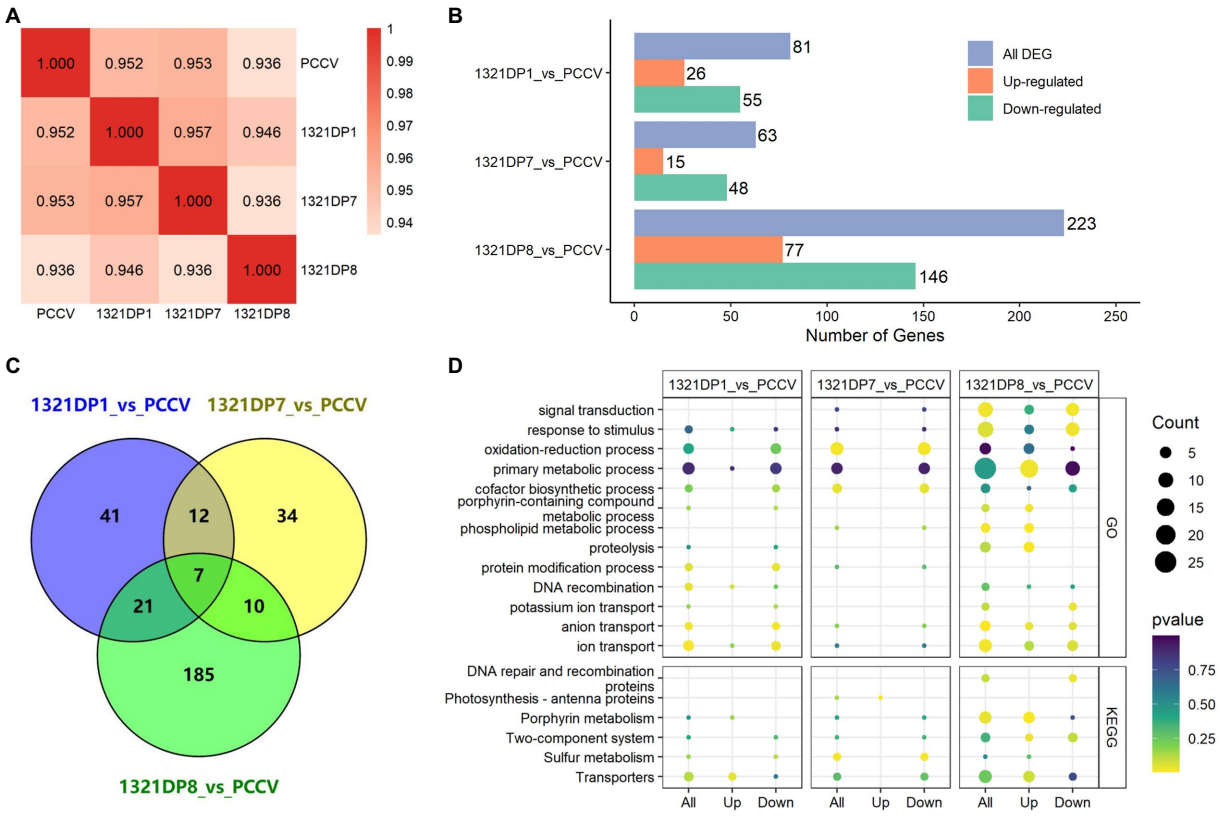

FIGURE 2
RNA-seq results analysis. **(A)** Pearson correlation analysis between samples. **(B)** Number of DEGs in three mutants. **(C)** Venn diagram of the DEGs in three mutants. **(D)** Pathway enrichment analysis results using DEGs.

Based on the fold changes, we identified the top 20 expressed genes under ethanol-treated conditions of 1321DP8 (Table 4). The top 20 up-regulated expressed genes were mostly hypothetical proteins (7), transcript factors (TFs) (5), transporters (2), recombinase family protein, magnesium-protoporphyrin IX monomethyl ester (oxidative) cyclase, heme oxygenase (biliverdin-producing), acyltransferase family protein, tetratricopeptide repeat protein, and tRNA-Val. Several TFs or TF related genes such as HATPase_c (*sll0790*), Resolvase (*slr8029*), cNMP-binding (*slr0607*), Radical_SAM (*sll1876*), and DUF559 (*SGL_RS19135*).

Top down-regulated genes included hypothetical proteins (9), TF or TF domain-containing proteins (DUF6, DUF45, DUF4269), DnaJ family molecular chaperone (*sll0909*), DNA cytosine methyltransferase, putative toxin-antitoxin system toxin component, TMEM family proteins (*slr1170* and *slr0305*), transposase and CRISPR-associated endonuclease Cas1 (*slr7071*). Compared with our previous OMIC studies, these TOP genes were not highlighted before (Wang et al., 2012; Zhu et al., 2015). Interestingly, we also found that most hypothetical proteins were on the top expressed gene list, suggesting their possible essential unknown protection functions (Wang et al., 2012).

DnaJ and DnaK belong to the heat shock protein family and are common stress responders. DnaJ family molecular chaperone (*sll0909*) was highly down-regulated in DP8 cells under ethanol stress. Similarly, we found a heat-shock DnaK homolog (*slr0086*) induced by ethanol stress (Wang et al., 2012). In this case, DP8 proteins may compensate for the defensive functions of DnaJ family proteins, at least at the transcriptional level.

It is well known that transcription factors (TFs) are involved in biofuel tolerance in bacteria (Ishihama, 2010; Fu et al., 2016). In cyanobacteria, a transcriptional regulator, *sll0794*, was involved in ethanol resistance directly, and the potential regulatory network mediated by *sll0794* was also proposed (Song et al., 2014). Through deletion mutagenesis, four TF, *slr0724*, *sll1392*, *sll1712*, and *sll0690*, were proven to be responsible for ethanol (Wang et al., 2012; Zhu et al., 2015) or butanol (Zhu et al., 2013) stress, respectively. In this study, unlike TFs mentioned above (*sll0794*, *slr0724*, *sll1392*, *sll1712*, and *sll0690*), significantly expressed novel TFs or TF-related genes, especially those TFs with unknown functions (DUF), were observed in 1321DP8, indicating CAHS might regulate biofuel stresses using different complicated pathways.

Strikingly, a toxin-antitoxin (TA) *ssl7046* was also highly down-regulated in DP8 cells. With no exactly known function, *ssl7046* was so far annotated as a ribonuclease. TA systems are involved in bacterial stress adaptation by regulating cell growth or death. They are abundantly existed in cyanobacteria but with little understanding (Fei et al., 2018). This is also the first report about TA systems involved in cyanobacteria against ethanol stress. Interestingly, CRISPR-associated endonuclease Cas1 (*slr7071*; Hein

10.3389/fmicb.2022.1091502

TABLE 4 Top20 up-regulated and down-regulated DEGs in 1321DP8.

Regulation	Gene ID	log2FoldChange	Value of *p*	Gene description	TF family
Up-regulated	*slr5013*	6.948	0.000	Hypothetical protein	
	slr8044	5.862	0.023	Hypothetical protein	
	slr0951	5.862	0.023	2-C-methyl-D-erythritol 4-phosphate cytidylyltransferase	
	sll0790	5.644	0.041	HAMP domain-containing histidine kinase	HATPase_c
	ST6803t05	3.075	0.035	tRNA-Val	
	sll1874	2.912	0.000	Magnesium-protoporphyrin IX monomethyl ester (oxidative) cyclase	
	sll1875	2.849	0.000	Heme oxygenase (biliverdin-producing)	
	slr1667	2.696	0.000	Hypothetical protein	
	sll1192	2.545	0.003	Fluoride efflux transporter CrcB	
	slr2103	2.238	0.000	Acyltransferase Family protein	
	SGL_RS20165	2.234	0.030	Hypothetical protein	
	sll1160	2.128	0.044	Hypothetical protein	
	slr1920	1.975	0.014	Hypothetical protein	
	slr8029	1.958	0.025	Recombinase family protein	Resolvase
	slr0607	1.933	0.004	Cyclic nucleotide-binding domain-containing protein	cNMP_binding
	slr1052	1.908	0.001	Tetratricopeptide repeat protein	
	slr1452	1.895	0.000	Sulfate ABC transporter substrate-binding protein	
	sll0419	1.880	0.033	Hypothetical protein	
	sll1876	1.818	0.000	Oxygen-independent coproporphyrinogen III oxidase	Radical_SAM
	SGL_RS19135	1.806	0.000	DUF559 domain-containing protein	DUF559
Down-regulated	*slr5082*	−6.447	0.002	Rpn family recombination-promoting nuclease/putative transposase	
	sll5062	−6.447	0.002	Hypothetical protein	
	slr6103	−6.161	0.007	Protein of unknown function DUF45	DUF45
	sll1005	−6.161	0.007	Nucleoside triphosphate pyrophosphohydrolase	
	slr1148	−6.161	0.007	DUF4269 domain-containing protein	DUF4269
	sll1319	−5.804	0.023	DMT family transporter	DUF6
	SGL_RS20130	−5.804	0.023	Hypothetical protein	
	SGL_RS19080	−5.586	0.041	Hypothetical protein	
	SGL_RS19680	−5.586	0.041	Hypothetical protein	
	SGL_RS19235	−5.586	0.041	Hypothetical protein	
	sll0909	−4.032	0.000	DnaJ family molecular chaperone	

(Continued)

TABLE 4 (Continued)

Regulation	Gene ID	log2FoldChange	Value of *p*	Gene description	TF family
	SGL_RS20145	−3.375	0.008	Hypothetical protein	
	sll0494	−3.375	0.008	Hypothetical protein	
	SGL_RS19495	−3.328	0.000	Hypothetical protein	
	slr6031	−3.251	0.013	DNA cytosine methyltransferase	
	ssl7046	−3.115	0.022	Putative toxin-antitoxin system toxin component	
	sll8033	−3.115	0.022	Hypothetical protein	
	slr1170	−2.840	0.000	TMEM165/GDT1 family protein	
	slr0305	−2.676	0.012	TVP38/TMEM64 family protein	
	slr7071	−2.570	0.018	CRISPR-associated endonuclease Cas1	

et al., 2013) was also showed up on the TOP down-regulated gene list. The links between the TA system and CRISPR are unclear; however, the expression of DP8 decreases the transcriptional levels of both TA and CRISPR systems in *Synechocystis*. Further investigations would be required to explore their connections.

4. Conclusion

In conclusion, we reported the hetero-expression of three CAHS proteins in a model cyanobacterium *Synechocystis*. TDP-expressed *Synechocystis* showed sensitive and tolerant phenotypes under different abiotic stresses. Notably, this analysis revealed that the induction of heat-shock protein and transporters, and modification of TCS, TA, and CRISPR systems were the central protection mechanisms against biofuel stress. The analysis provided a novel direction for engineering ethanol tolerance in cyanobacterium *Synechocystis*.

Data availability statement

The datasets presented in this study can be found in online repositories. The names of the repository/repositories and accession number(s) can be found at: https://www.ncbi.nlm.nih.gov/, PRJNA888863.

Author contributions

HZ and QYL performed the experiments and wrote the initial manuscript. QYL and BW designed the study. ZC and JW analyzed the data and revised the manuscript. All authors have read and approved the final manuscript.

Funding

This research was supported by grants from the National Key R&D Program of China (2021YFA0910800, 2020YFA0908703, and 2018YFA0902500), and the National Natural Science Foundation of China (41876188).

Conflict of interest

QL and BW were employed by Shenzhen Link Spider Technology Co., Ltd.

The remaining authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Supplementary material

The Supplementary material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fmicb.2022.1091502/full#supplementary-material

References

Agostoni, M., Logan-Jackson, A. R., Heinz, E. R., Severin, G. B., Bruger, E. L., Waters, C. M., et al. (2018). Homeostasis of second messenger cyclic-di-AMP is critical for Cyanobacterial fitness and acclimation to abiotic stress. *Front. Microbiol.* 9:1121. doi: 10.3389/fmicb.2018.01121

Arakawa, K., and Numata, K. (2021). Reconsidering the "glass transition" hypothesis of intrinsically unstructured CAHS proteins in desiccation tolerance of tardigrades. *Mol. Cell* 81, 409–410. doi: 10.1016/j.molcel.2020.12.007

Bi, Y., Pei, G., Sun, T., Chen, Z., Chen, L., and Zhang, W. (2018). Regulation mechanism mediated by trans-encoded sRNA Nc117 in short chain alcohols tolerance in *Synechocystis sp.* PCC 6803. *Front. Microbiol.* 9:863. doi: 10.3389/fmicb.2018.00863

Boothby, T. C. (2021). Water content influences the vitrified properties of CAHS proteins. *Mol. Cell* 81, 411–413. doi: 10.1016/j.molcel.2020.12.009

Boothby, T. C., Tapia, H., Brozena, A. H., Piszkiewicz, S., Smith, A. E., Giovannini, I., et al. (2017). Tardigrades use intrinsically disordered proteins to survive desiccation. *Mol. Cell* 65, 975–984.e5. doi: 10.1016/j.molcel.2017.02.018

Chavali, S., Gunnarsson, A., and Babu, M. M. (2017). Intrinsically disordered proteins adaptively reorganize cellular matter during stress. *Trends Biochem. Sci.* 42, 410–412. doi: 10.1016/j.tibs.2017.04.007

Chen, L., Wu, L., Wang, J., and Zhang, W. (2014). Butanol tolerance regulated by a two-component response regulator Slr1037 in photosynthetic *Synechocystis sp.* PCC 6803. *Biotechnol. Biofuels* 7:89. doi: 10.1186/1754-6834-7-89

Crilly, C. J., Brom, J. A., Warmuth, O., Esterly, H. J., and Pielak, G. J. (2022). Protection by desiccation-tolerance proteins probed at the residue level. *Protein Sci.* 31, 396–406. doi: 10.1002/pro.4231

Fei, Q., Gao, E. B., Liu, B., Wei, Y., and Ning, D. (2018). A toxin-antitoxin system VapBC15 from *Synechocystis sp.* PCC 6803 shows distinct regulatory features. *Genes (Basel)* 9:173. doi: 10.3390/genes9040173

Fu, Y., Chen, L., and Zhang, W. (2016). Regulatory mechanisms related to biofuel tolerance in producing microbes. *J. Appl. Microbiol.* 121, 320–332. doi: 10.1111/jam.13162

Hagemann, M., Schoor, A., Jeanjean, R., Zuther, E., and Joset, F. (1997). The stpA gene form *Synechocystis sp.* strain PCC 6803 encodes the glucosylglycerol-phosphate phosphatase involved in cyanobacterial osmotic response to salt shock. *J. Bacteriol.* 179, 1727–1733. doi: 10.1128/jb.179.5.1727-1733.1997

Hein, S., Scholz, I., Voß, B., and Hess, W. R. (2013). Adaptation and modification of three CRISPR loci in two closely related cyanobacteria. *RNA Biol.* 10, 852–864. doi: 10.4161/rna.24160

Hesgrove, C., and Boothby, T. C. (2020). The biology of tardigrade disordered proteins in extreme stress tolerance. *Cell Commun. Signal* 18:178. doi: 10.1186/s12964-020-00670-2

Ishihama, A. (2010). Prokaryotic genome regulation: multifactor promoters, multitarget regulators and hierarchic networks. *FEMS Microbiol. Rev.* 34, 628–645. doi: 10.1111/j.1574-6976.2010.00227.x

Janis, B., Belott, C., and Menze, M. A. (2018). Role of intrinsic disorder in animal desiccation tolerance. *Proteomics* 18:e1800067. doi: 10.1002/pmic.201800067

Koshland, D., and Tapia, H. (2019). Desiccation tolerance: an unusual window into stress biology. *Mol. Biol. Cell* 30, 737–741. doi: 10.1091/mbc.E17-04-0257

Lamb, J. J., and Hohmann-Marriott, M. F. (2017). Manganese acquisition is facilitated by PilA in the cyanobacterium *Synechocystis sp.* PCC 6803. *PLoS One* 12:e0184685. doi: 10.1371/journal.pone.0184685

Liu, J., Chen, L., Wang, J., Qiao, J., and Zhang, W. (2012). Proteomic analysis reveals resistance mechanism against biofuel hexane in *Synechocystis sp.* PCC 6803. *Biotechnol. Biofuels* 5:68. doi: 10.1186/1754-6834-5-68

Liu, D., and Pakrasi, H. B. (2018). Exploring native genetic elements as plug-in tools for synthetic biology in the cyanobacterium *Synechocystis sp.* PCC 6803. *Microb. Cell Factories* 17:48. doi: 10.1186/s12934-018-0897-8

Machado, I. M., and Atsumi, S. (2012). Cyanobacterial biofuel production. *J. Biotechnol.* 162, 50–56. doi: 10.1016/j.jbiotec.2012.03.005

Malki, A., Teulon, J. M., Camacho-Zarco, A. R., Chen, S. W., Adamski, W., Maurin, D., et al. (2022). Intrinsically disordered Tardigrade proteins self-assemble into fibrous gels in response to environmental stress. *Angew. Chem. Int. Ed. Engl.* 61:e202109961. doi: 10.1002/anie.202109961

Mikkat, S., Effmert, U., and Hagemann, M. (1997). Uptake and use of the osmoprotective compounds trehalose, glucosylglycerol, and sucrose by the cyanobacterium *Synechocystis sp.* PCC6803. *Arch. Microbiol.* 167, 112–118. doi: 10.1007/s002030050423

Pei, G., Chen, L., Wang, J., Qiao, J., and Zhang, W. (2014). Protein network signatures associated with exogenous biofuels treatments in Cyanobacterium *Synechocystis sp.* PCC 6803. *Front. Bioeng. Biotechnol.* 2:48. doi: 10.3389/fbioe.2014.00048

Qiao, J., Wang, J., Chen, L., Tian, X., Huang, S., Ren, X., et al. (2012). Quantitative iTRAQ LC-MS/MS proteomics reveals metabolic responses to biofuel ethanol in cyanobacterial *Synechocystis sp.* PCC 6803. *J. Proteome Res.* 11, 5286–5300. doi: 10.1021/pr300504w

Reed, R. H., and Stewart, W. D. P. (1985). Osmotic adjustment and organic solute accumulation in unicellular cyanobacteria from freshwater and marine habitats. *Mar. Biol.* 88, 1–9. doi: 10.1007/BF00393037

Song, Z., Chen, L., Wang, J., Lu, Y., Jiang, W., and Zhang, W. (2014). A transcriptional regulator Sll0794 regulates tolerance to biofuel ethanol in photosynthetic *Synechocystis sp.* PCC 6803. *Mol. Cell. Proteomics* 13, 3519–3532. doi: 10.1074/mcp.M113.035675

Tanaka, A., Nakano, T., Watanabe, K., Masuda, K., Honda, G., Kamata, S., et al. (2022). Stress-dependent cell stiffening by tardigrade tolerance proteins that reversibly form a filamentous network and gel. *PLoS Biol.* 20:e3001780. doi: 10.1371/journal.pbio.3001780

Tanaka, S., Tanaka, J., Miwa, Y., Horikawa, D. D., Katayama, T., Arakawa, K., et al. (2015). Novel mitochondria-targeted heat-soluble proteins identified in the anhydrobiotic Tardigrade improve osmotic tolerance of human cells. *PLoS One* 10:e0118272. doi: 10.1371/journal.pone.0118272

Tian, X., Chen, L., Wang, J., Qiao, J., and Zhang, W. (2013). Quantitative proteomics reveals dynamic responses of *Synechocystis sp.* PCC 6803 to next-generation biofuel butanol. *J. Proteome* 78, 326–345. doi: 10.1016/j.jprot.2012.10.002

Wang, J., Chen, L., Huang, S., Liu, J., Ren, X., Tian, X., et al. (2012). RNA-seq based identification and mutant validation of gene targets related to ethanol resistance in cyanobacterial *Synechocystis sp.* PCC 6803. *Biotechnol. Biofuels* 5:89. doi: 10.1186/1754-6834-5-89

Wang, Y., Sun, T., Gao, X., Shi, M., Wu, L., Chen, L., et al. (2016). Biosynthesis of platform chemical 3-hydroxypropionic acid (3-HP) directly from CO2 in cyanobacterium *Synechocystis sp.* PCC 6803. *Metab. Eng.* 34, 60–70. doi: 10.1016/j.ymben.2015.10.008

Wu, W., Du, W., Gallego, R. P., Hellingwerf, K. J., van der Woude, A. D., and Branco Dos Santos, F. (2020). Using osmotic stress to stabilize mannitol production in *Synechocystis sp.* PCC6803. *Biotechnol. Biofuels* 13:117. doi: 10.1186/s13068-020-01755-3

Yagi-Utsumi, M., Aoki, K., Watanabe, H., Song, C., Nishimura, S., Satoh, T., et al. (2021). Desiccation-induced fibrous condensation of CAHS protein from an anhydrobiotic tardigrade. *Sci. Rep.* 11:21328. doi: 10.1038/s41598-021-00724-6

Zhu, Y., Pei, G., Niu, X., Shi, M., Zhang, M., Chen, L., et al. (2015). Metabolomic analysis reveals functional overlapping of three signal transduction proteins in regulating ethanol tolerance in cyanobacterium *Synechocystis sp.* PCC 6803. *Mol. BioSyst.* 11, 770–782. doi: 10.1039/c4mb00651h

Zhu, H., Ren, X., Wang, J., Song, Z., Shi, M., Qiao, J., et al. (2013). Integrated OMICS guided engineering of biofuel butanol-tolerance in photosynthetic *Synechocystis sp.* PCC 6803. *Biotechnol. Biofuels* 6:106. doi: 10.1186/1754-6834-6-106

frontiers | Frontiers in Microbiology

TYPE Original Research
PUBLISHED 02 February 2023
DOI 10.3389/fmicb.2023.1123081

OPEN ACCESS

EDITED BY
Xiaoming Tan,
Hubei University,
China

REVIEWED BY
Jiangxin Wang,
Shenzhen University,
China
Stephan Klähn,
Helmholtz Centre for Environmental Research,
Helmholtz Association of German Research Centres (HZ), Germany

*CORRESPONDENCE
Tao Sun
✉ tsun@tju.edu.cn
Lei Chen
✉ lchen@tju.edu.cn

SPECIALTY SECTION
This article was submitted to Microbiotechnology, a section of the journal Frontiers in Microbiology

RECEIVED 13 December 2022
ACCEPTED 17 January 2023
PUBLISHED 02 February 2023

CITATION
Dong Z, Sun T, Zhang W and Chen L (2023) Improved salt tolerance of *Synechococcus elongatus* PCC 7942 by heterologous synthesis of compatible solute ectoine. *Front. Microbiol.* 14:1123081. doi: 10.3389/fmicb.2023.1123081

Improved salt tolerance of *Synechococcus elongatus* PCC 7942 by heterologous synthesis of compatible solute ectoine

Zhengxin Dong[1,2], Tao Sun[1,2,3]*, Weiwen Zhang[1,2,3] and Lei Chen[1,2]*

[1]Laboratory of Synthetic Microbiology, School of Chemical Engineering and Technology, Tianjin University, Tianjin, China, [2]Frontier Science Center for Synthetic Biology and Key Laboratory of Systems Bioengineering, Ministry of Education of China, Tianjin, China, [3]Center for Biosafety Research and Strategy, Tianjin University, Tianjin, China

Salt stress is one of the essential abiotic stresses for the survival of cyanobacteria. However, the realization of large-scale cultivation of cyanobacteria is inseparable from the utilization of abundant seawater resources. Therefore, research on the regulatory mechanism, as well as the improvement of salt tolerance of cyanobacteria is fundamental. Ectoine, a compatible solute which was found in halophilic microorganisms, has potentiality to confer salt tolerance. Here in this article, the salt tolerance of *Synechococcus elongatus* PCC 7942 (Syn7942) was significantly improved *via* expressing the ectoine biosynthetic pathway, reaching an increased final OD_{750} by 20% under 300mM NaCl and 80% under 400mM NaCl than that of wild-type (WT), respectively. Encouragingly, the engineered strain could even survive under 500mM NaCl which was lethal to WT. In addition, by introducing the ectoine synthetic pathway into the sucrose-deficient strain, the salt tolerance of the obtained strain Syn7942/Δsps-ect was restored to the level of WT under 300mM NaCl stress, demonstrating that ectoine could substitute for sucrose to combat against salt stress in Syn7942. In order to study the difference in the regulation of mechanism on the salt adaptation process after replacing sucrose with ectoine, transcriptomic analysis was performed for Syn7942/Δsps-ect and WT. The differentially expressed gene analysis successfully identified 19 up-regulated genes and 39 down-regulated genes in Syn7942/Δsps-ect compared with WT under salt stress condition. The results also showed that the global regulation of Syn7942/Δsps-ect and WT had certain differences in the process of salt adaptation, in which Syn7942/Δsps-ect reduced the demand for the intensity of sulfur metabolism in this process. This study provides a valuable reference for further salt tolerance engineering in cyanobacteria.

KEYWORDS

cyanobacteria, salt stress, compatible solutes, ectoine, transcriptome

Introduction

Cyanobacteria can use solar energy and CO_2 in the air to synthesize organic compounds, realizing the negative carbon economy in the process, which have received widespread attention as autotrophic cell factories (Kato et al., 2022; Liu et al., 2022; Tan et al., 2022). As a model cyanobacterium, *Synechococcus elongatus* PCC 7942 (hereafter Syn7942), is amenable to genetic manipulation with the development of toolboxes and has been employed as a biological chassis for chemical production (Kim et al., 2017; Sun et al., 2018; Sengupta et al., 2019; Zhang M. et al., 2022). Nowadays, it is expected to carry out industrialized production of biofuels and chemicals through the large-scale cultivation of cyanobacteria (Farrokh et al., 2019; Davies et al., 2021; Wang et al., 2021), which requires the employment of seawater resources with rich reserves (Cui et al., 2020). Therefore, research on

improving the salt tolerance of cyanobacteria is required to enable their cultivation in seawater resources to produce high value-added chemicals. Moreover, as a model freshwater cyanobacterium, the engineering and regulation mechanism will provide a valuable reference for further salt tolerance engineering in cyanobacteria.

Most of halophilic microorganisms accumulate organic compatible solutes to maintain the balance of osmotic pressure of the intracellular and the external environment (Gunde-Cimerman et al., 2018). Among them, ectoine was initially found in *Ectothiorhodospira halochloris* (Galinski et al., 1985) and is widely used in the fields of food, cosmetics, and medicine (Liu et al., 2021; Kauth and Trusova, 2022; Zhang H. et al., 2022). Recent studies have shown that there was also the native synthesis of ectoine in some microalgae (Fenizia et al., 2020). The biosynthetic pathway of ectoine has been analyzed, and its heterologous synthesis has been achieved in both *Escherichia coli* and *Corynebacterium glutamicum* (Giesselmann et al., 2019; Chen J. et al., 2020; Zhang H. et al., 2022). The synthesis of ectoine depends on gene *ectB* encoding L-2,4-diaminobutyrate transaminase, *ectA* encoding 2,4-diaminobutyrate acetyltransferase, *ectC* encoding ectoine synthase, and using L-aspartate-β-semialdehyde as substrate (Göller et al., 1998; Calderón et al., 2004; Schwibbert et al., 2011). However, the only compatible solute in Syn7942 is sucrose (Klähn and Hagemann, 2011). Heterologous synthesis compatible solute to improve salt tolerance has been considered a very effective method (Waditee-Sirisattha et al., 2012; Singh et al., 2013; Cui et al., 2021). For example, the heterologous synthesis of glucosylglycerol in *Synechococcus elongatus* UTEX 2973 resulted in a 62% increase in growth under 0.5 M NaCl conditions (Cui et al., 2021). Therefore, it is feasible to synthesize ectoine heterologously to increase salt tolerance in Syn7942.

In this study, aiming to engineering the salt tolerance of the model cyanobacterium Syn7942, the first study of heterologous synthesis of ectoine in cyanobacteria was achieved, and the salt tolerance of Syn7942 was successfully improved. This provides a potential solution for improving the salt tolerance of cyanobacteria while producing high value-added products. Then, the recovery of salt tolerance of the Syn7942/Δsps-ect strain was achieved by introducing the ectoine synthesis pathway into the sucrose synthesis deficient strain (Syn7942/Δsps), indicating that ectoine was able to substitute for sucrose to combat against salt stress in Syn7942. Finally, to deciphering the mechanism of improved salt tolerance in the engineered Syn7942/Δsps-ect, the transcriptional differences between strains Syn7942/Δsps-ect and WT under salt stress condition were further studied by comparative transcriptomics, and the differentially expressed genes (DEGs) were identified. This study provided valuable information for engineering salt-tolerant cyanobacteria.

Materials and methods

Bacterial growth conditions and salt-stress treatment

The WT and engineered Syn7942 were grown at 37°C in BG11 liquid medium (pH 7.5) or on agar plates under a light intensity of approximately 100 μmol photons m^{-2} s^{-1} in an HNY-211B Illuminating incubator Shaker of 200 rpm in 100 ml flask (Honour, Tianjin, China). Twenty-five milligrams per liter of chloramphenicol (Solarbio, Beijing, China) was added during the culture of engineering strains. Different salt concentrations of media used in the culture process were obtained by adding a suitable amount of 3 M to 0 M NaCl BG11 medium. Cell density was measured at 750 nm (OD_{750}) by an ELx808 Absorbance Microplate Reader (BioTek, VT, United States). *E. coli* TOP10 was grown in LB liquid medium or LB agar plates, with 50 mg/l of chloramphenicol or spectinomycin for screening and maintaining the stability of engineered bacteria.

Construction of strains and plasmids

The strains used in this study are listed in Table 1. Among them, *E. coli* TOP10 was used for plasmid construction. pSI-ect vector with a chloramphenicol-resistant cassette was constructed for express ectoine production cassette based on laboratory plasmid (Li et al., 2018). To knock out the sucrose production pathway and express the ectoine synthesis pathway after the knockout of the sucrose production pathway, pSPS-ect was constructed by replacing the upstream and downstream homologous arms based on pSI-ect, and pSPS was built by deleting an ectoine expression cassette based on pSPS-ect. Primers and plasmids used for this study are listed in Supplementary Table S1. All primers were synthesized by Azenta (Suzhou, China). The genes *ectA*, *ectB*, and *ectC* were synthesized by Azenta based on the *Halomonas elongata* after codon optimization. The sequences are shown in Supplementary Table S2. The template plasmids were purified by FastPure Plasmid Mini Kit (Vazyme Biotech, Nanjing, China). The target fragments were amplified by Phanta Super-Fidelity DNA Polymerase (Vazyme Biotech, Nanjing, China) and purified by FastPure Gel DNA Extraction Mini Kit (Vazyme Biotech, Nanjing, China). Then, fragments were ligated by ClonExpress MultiS One Step Cloning Kit (Vazyme Biotech, Nanjing, China) or Golden Gate (Thermo Fisher Scientific Inc., CA, United States). All constructs were verified by PCR and Sanger sequencing. After the constructed plasmids were extracted and cut through NdeI (Thermo Fisher Scientific Inc., CA, United States), and were transformed into Syn7942 by the method of natural transformation.

Ectoine extraction and measurement

One milligram liter of cultures on the 7th day was taken out and placed at −80°C. When preparing the samples, the cultures were taken out from −80°C, placed at 70°C for 2 h, centrifuged at 13,000 rpm. And the supernatant was taken for ectoine concentration determination. Based on the previous study (Ning et al., 2016), the ectoine was detected using high-performance liquid chromatography (Agilent 1,260 Series HPLC, Agilent Technologies, Santa Clara, CA, United States), and the concentration of ectoine in the samples was determined using the ectoine standard (Sigma-Aldrich, Shanghai, China). The chromatographic column was Ultimate AQ-C18, 5 μm 4.6*250 mm (Welch, China), and the mobile phase condition was ultrapure water of 0.6 ml/min. The ectoine was monitored using a UV detector at a wavelength of 210 nm.

Transcriptomic analysis

WT and Syn7942/Δsps-ect were cultured in BG11 medium with 0 mM as control-1 and control-2 and 300 mM NaCl as experimental-1

TABLE 1 Strains used in this study.

Strains	Genotype or relevant features	References
WT	*Synechococcus elongatus* PCC 7942	Laboratory storage
Syn7942/NSI-ect	NSI:P_{cpc560}-ectABC-T_{rbcL}; cm^R	In this study
Syn7942/Δsps	Synpcc7942_0808:Pcat-cm^R-T_{rrnB};cmR	In this study
Syn7942/Δsps-ect	Synpcc7942_0808:P_{cpc560}-ectABC-T_{rbcL}	In this study

and experimental-2, respectively. On the fourth day, samples were collected and sent to Azenta for transcriptome sequencing and data analysis. Each sample had three biological replicates. The $\log_2$(fold-change) > 1.5 and the *value of p* <0.05 were set as the threshold for DEGs identification. KEGG enrichment analysis was done using TBtools (Version v1.098769; Chen C. et al., 2020). There was still a small number of reads for the *Synpcc7942_0808* gene in Syn7942/Δsps-ect. In order to exclude data interference, the number of reads for the *Synpcc7942_0808* gene in Syn7942/Δsps-ect was ignored.

The transcriptome data has been uploaded to GEO database (GSE222067).

Results and discussion

Improved salt tolerance of Syn7942 by heterologous expression of the ectoine biosynthetic pathway

The gene cluster *ectABC* was codon-optimized for Syn7942 and then has been chemically synthesized and introduced into the NSI site of Syn7942 afterwards. The expression is controlled by the strong promoter P_{cpc560} (Figure 1). This engineered strain was named Syn7942/NSI-ect. To evaluate the salt tolerance of Syn7942/NSI-ect, WT was selected as the control strain, and the growth curves of Syn7942/NSI-ect and WT were measured in BG11 medium with addition of 0, 100, 200, 300, 400, and 500 mM NaCl, respectively. Under the growth condition of 0 mM NaCl, the OD_{750} of Syn7942/NSI-ect was decreased by 19% compared with that of WT on the 7th day, which may be caused by the competition between ectoine synthesis pathway and endogenous amino acid metabolism pathway for L-aspartate-4-semialdehyde. And the growth of Syn7942/NSI-ect and WT was almost the same under 100 and 200 mM NaCl conditions, while the OD_{750} of Syn7942/NSI-ect on the 7th day was increased by 20% than that of WT under 300 mM NaCl and increased by 80% under the condition of 400 mM NaCl (Figure 2A; Supplementary Figures S1A,B), demonstrating the improved salt tolerance of strain Syn7942/NSI-ect. Under 500 mM NaCl condition, the OD_{750} of Syn7942/NSI-ect on the 7th day could reach 0.362 while WT could not survive at all. This suggested that the salt tolerance of Syn7942 could be improved by the introduction of ectoine biosynthesis pathway.

The production of ectoine in Syn7942/NSI-ect under different salt concentrations on the 7th day was then determined using the external standard method. The standard curve was shown in Supplementary Figure S2. As shown in Figure 2B, under the condition of 0 mM NaCl, the yield of ectoine was 49.8 mg/l in Syn7942/NSI-ect. Moreover, under 100 mM NaCl, the yield of ectoine was significantly increased, reaching 98.9 mg/l. It has also been reported that ectoine production increased with the concentration of NaCl in the environment (Ning et al., 2016; Yu et al., 2022), and this phenomenon also occurred in the synthesis of glucosylglycerol and sucrose (Song et al., 2016; Cui et al., 2021). With the further increase of salt concentration, the yield of ectoine decreased compared with under 100 mM NaCl, most probably due to the worse growth caused by salt stress.

In previous studies, the salt environment was reported to be favorable for the synthesis of ectoine (Ning et al., 2016). Here, our results showed that the synthesis of ectoine could significantly improve the salt tolerance of Syn7942/NSI-ect, and the amount of ectoine synthesis was responsive to the high-salt environment.

Ectoine could substitute for sucrose to combat against salt stress in Syn7942

The accumulation of compatible solutes is an essential way for microorganisms to resist the high-salt environment, and the only compatible solute in Syn7942 is sucrose (Klähn and Hagemann, 2011). The loss of the sucrose synthesis pathway led to the generation of a salt-sensitive strain, which indicated that sucrose was the only compatible solute. However, sucrose is not considered an ideal osmocompatible substance against salt stress (Klähn and Hagemann, 2011). In a study of adaptations to salt tolerance in *Synechocystis* sp. PCC 6803, the accumulation of glucosylglycerol was much higher than sucrose, and the contribution of glucosylglycerol was much higher than that of sucrose during the salt adaptation process; therefore, in natural selection, sucrose was more critical as an energy storage substance rather than playing a role in salt tolerance (Klähn et al., 2021). As a compatible solute retained by natural evolution in halophilic bacteria (Göller et al., 1998), ectoine was believed to have unique advantages in salt tolerance. Here, to clarify the role of ectoine and exclude the influence of the native compatible solute sucrose in salt stress adaption in Syn7942, the sucrose production pathway of Syn7942 was knocked out to obtain strain Syn7942/Δsps. The growth of Syn7942/Δsps was measured under the conditions of 0, 100, 200, 300, 400, and 500 mM NaCl (Supplementary Figure S1C), and the results showed that the OD_{750} of Syn7942/Δsps on the 7th day was decreased to 62% under the 100 mM NaCl condition, 36% under the 200 mM NaCl condition compared with WT, and the strain could not survive under the 300 mM NaCl condition (Figure 3A).

Next, the ectoine synthesis module was introduced into Syn7942 instead of sucrose synthesis to obtain strain Syn7942/Δsps-ect; and the salt-adaptive ability was then investigated for growth at 0, 100, 200, 300, 400, and 500 mM NaCl, respectively (Supplementary Figure S1D). As shown in Figures 3B,C, compared with Syn7942/Δsps, the introduction of the ectoine synthesis module successfully recovered the salt tolerance of Syn7942/Δsps to WT level under 100, 200 and 300 mM NaCl conditions. Ectoine production was also determined in Syn7942/Δsps-ect grown under different salt concentrations. As shown in Figure 3D, the yield of ectoine on the 7th day was 53 mg/l under 0 mM NaCl, the similar level as that of Syn7942/NSI-ect. The yields of ectoine in Syn7942/Δsps-ect grown under 100, 200, and 300 mM NaCl conditions were 115.2 mg/L, 112.5 mg/L, and 72.9 mg/L, respectively, which were slightly improved compared with the yields in Syn7942/NSI-ect under the same conditions, probably due to that an additional part of ectoine need to be synthesized to make up for the lack of sucrose under the condition of salt stress. In addition, in Syn7942/Δsps-ect the lack of sucrose accumulation saved a part of carbon source for ectoine synthesis.

Global transcriptomic analysis

To explore the working mechanism of ectoine in salt stress adaptation in Syn7942/Δsps-ect, comparative transcriptomic analysis was performed in Syn7942/Δsps-ect and WT, under 0 and 300 mM NaCl conditions, respectively. After data quality control, more than 16 million reads per sample were mapped to the reference genome, and the detected genes in each sample covered more than 94% of all the predicted genes (Supplementary Table S3).

As shows in Figure 4, under 0 mM NaCl condition, only 11 DEGs were identified between Syn7942/Δsps-ect and WT. This indicated that under the condition of 0 mM NaCl, ectoine synthesis did not have significant effect on the overall metabolism.

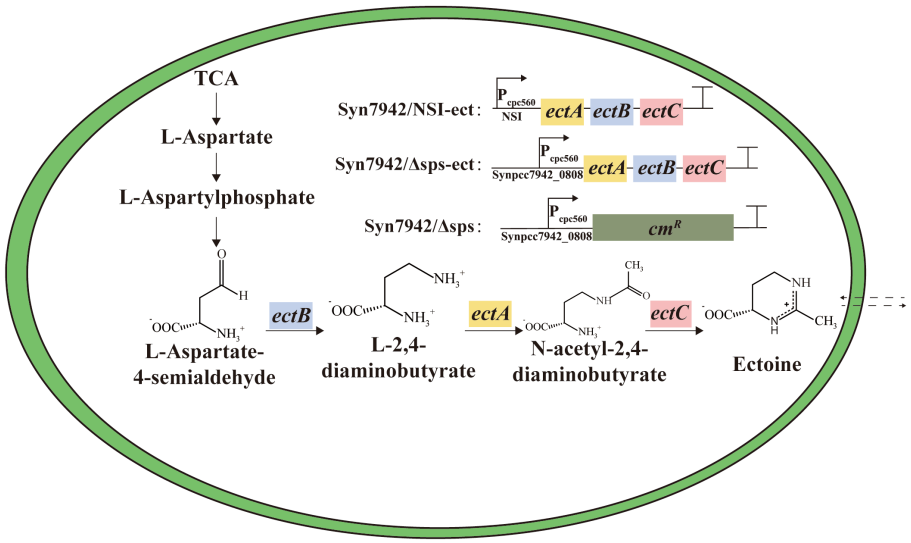

FIGURE 1
The reconstructed ectoine biosynthetic pathway in Syn7942. Syn7942/NSI-ect: indicates that the expression cassette indicated in the legend was inserted at the neutral site I (NSI); Syn7942/Δsps and Syn7942/Δsps-ect: indicate that the endogenous *Synpcc7942_0808* gene was replaced using the expression cassette in the legend; TCA: indicates tricarboxylic acid cycle.

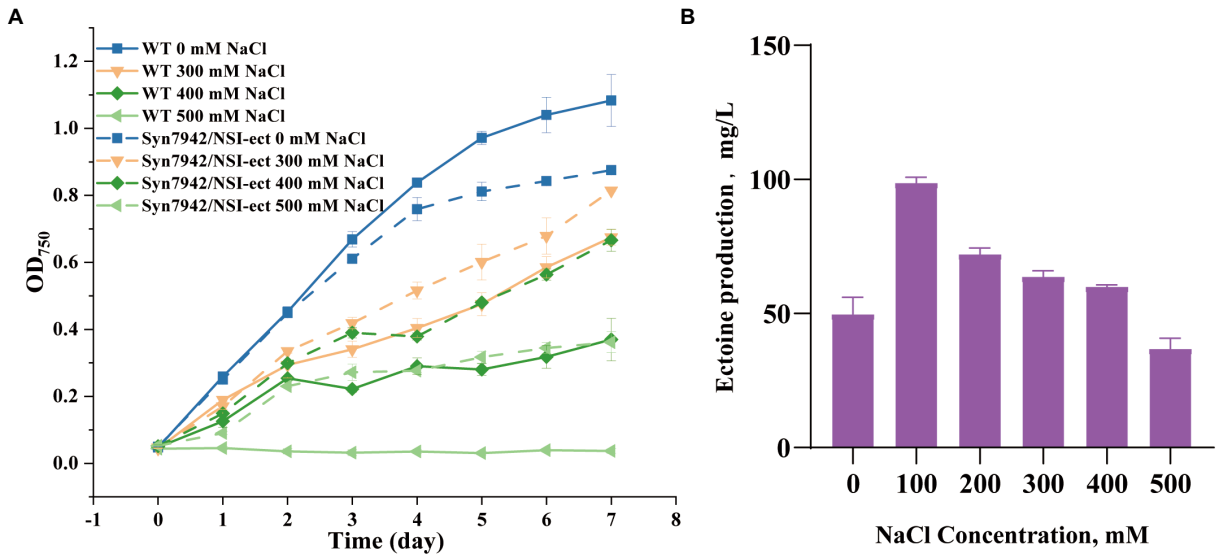

FIGURE 2
Growth curves and ectoine production of WT and Syn7942/NSI-ect. **(A)** Growth curves of WT and Syn7942/NSI-ect under 0, 300, 400, 500mM NaCl; **(B)** Ectoine production of Syn7942/NSI-ect under 0, 100, 200, 300, 400, 500mM NaCl.

Analysis of global regulation of salt adaptation in WT and Syn7942/Δsps-ect

In the WT, 246 genes were up-regulated and 331 down-regulated when grown in presence of 300 mM NaCl compared to the reference conditions (0 mM NaCl). The same comparison revealed 182 up-regulated and 287 down-regulated genes in strain Syn7942/Δsps-ect (Figure 4). Among them, 104 up-regulated and 198 down-regulated genes were shared between WT and Syn7942/Δsps-ect (Supplementary Figure S3), suggesting that these DEGs played important roles in the salt stress adaption process in both strains of Syn7942/Δsps-ect and WT.

KEGG pathway enrichment analysis was performed on DEGs of WT and Syn7942/Δsps-ect (Figure 5). Up-regulated genes of WT were mainly enriched in "nitrogen metabolism," "energy metabolism," "ABC transporters," "membrane transport," "O-antigen nucleotide sugar biosynthesis," "metabolism," "sulfur metabolism," "peptidoglycan biosynthesis and degradation proteins," "oxidative phosphorylation," "transporters," "photosynthesis proteins," "amino sugar and nucleotide sugar metabolism," and "biosynthesis of other secondary metabolites." Compared with up-regulated genes of WT, the number of enriched pathways of up-regulated genes of Syn7942/Δsps-ect was less, mainly enriched in "nitrogen metabolism," "energy metabolism," "exosome," "biosynthesis of other secondary metabolites," "glycolysis/gluconeogenesis" and "metabolism of terpenoids and polyketides."

Among them, "energy metabolism" and "nitrogen metabolism" were both enriched in the up-regulated genes of WT and Syn7942/Δsps-ect, but down-regulated under short-term salt stress (Billis

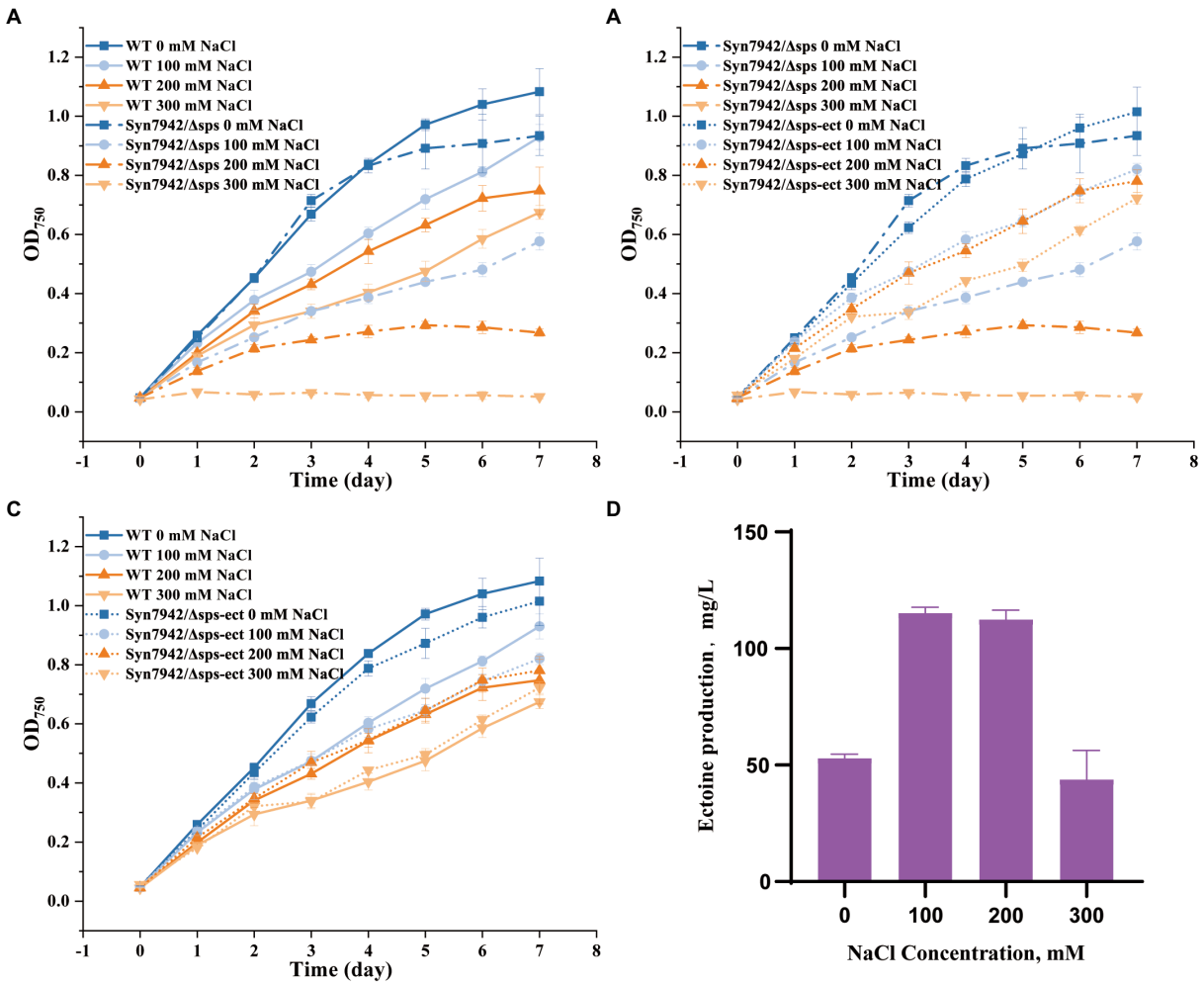

FIGURE 3
Growth curves and ectoine production of WT, Syn7942/Δsps, and Syn7942/Δsps-ect. **(A)** Growth curves of WT and Syn7942/Δsps under 0, 100, 200, 300mM NaCl; **(B)** Growth curves of Syn7942/Δsps and Syn7942/Δsps-ect under 0, 100, 200, 300mM NaCl; **(C)** Growth curves of WT and Syn7942/Δsps-ect under 0, 100, 200, 300mM NaCl; **(D)** Ectoine production of Syn7942/Δsps-ect under 0, 100, 200, 300mM NaCl.

et al., 2014). At the same time, in WT and Syn7942/Δsps-ect, genes *Synpcc7942_2016*, *Synpcc7942_2105* and *Synpcc7942_2107*, which encode nitrogen transporters, were significantly up-regulated, with an up-regulation fold of more than 8 folds. This showed that the enhancement of "energy metabolism" and "nitrogen metabolism" played important roles in helping the strains adapt to the high-salt environment for a long time. In a previous report, the operon Synpcc7942_2203 - Synpcc7942_2235 was unregulated in Syn7942 after 24 h salt stress (Billis et al., 2014). This phenomenon was not observed in this study, which may be caused by different salt stress times (in this experiment, strains were treated 4 days in 300 mM NaCl condition). The key gene of sucrose synthesis, *Synpcc7942_0808* encoding sucrose-phosphate synthase did not show significant change during salt stress ($\log_2$(fold-change = 0.83)). In addition, the "glycolysis/gluconeogenesis" pathway was enriched in the Syn7942/Δsps-ect. The enhancement of this pathway may facilitate ectoine production. It was reported that in the process of optimizing the biosynthesis of ectoine, increasing the carbon flux to ectoine is an important strategy (Ning et al., 2016; Giesselmann et al., 2019; Chen J. et al., 2020). The genes Synpcc7942_1312, Synpcc7942_1244 and Synpcc7942_0247 encoding ATP synthase were all up-regulated in WT and Syn7942/Δsps-ect, and the increase of ATP synthase expression was beneficial to the adaptation of cyanobacteria to high-salt environment (Soontharapirakkul et al., 2011).

The down-regulated genes of WT were mainly enriched in "chaperones and folding catalysts," "infectious disease: bacterial," "membrane trafficking," "mitochondrial biogenesis," "unclassified: genetic information processing," "human diseases," "two-component system," "signal transduction," "exosome" and "protein families: genetic information processing" (Figure 5B). Moreover, the down-regulated genes enrichment pathway analysis of Syn7942/Δsps-ect and WT were basically same (Figures 5B,D). The down-regulated genes of Syn7942/Δsps-ect were mainly enriched in "chaperones and folding catalysts," "membrane trafficking," "infectious disease: bacterial," "protein families: signaling and cellular processes," "mitochondrial biogenesis," "function unknown," "environmental information processing," "human diseases," "two-component system," "signal transduction" and "exosome" (Figure 5D). Genes related to signal transduction were down-regulated in WT and Syn7942/Δsps-ect, which was consistent with previous reports (Billis et al., 2014). This indicated that during the long-term salt adaptation process, the contribution of the signal transduction system to the strain's adaptation process under the high-salt environment decreased.

Analysis of differentially expressed genes in Syn7942/Δsps-ect compared to WT

In strain Syn7942/Δsps-ect, 19 genes were up-regulated and 39 down-regulated genes when grown in presence of 300 mM NaCl compared to the reference strains (WT). Among the DEGs, the gene *Synpcc7942_1531* encoding the molybdenum ABC transporter and the gene *Synpcc7942_1530* encoding the molybdenum-pterin binding domain were up-regulated to 4.6 and 3.9 folds, respectively. In bacteria, ingested molybdenum can combine with molybdenum cofactors, regulate the activity of molybdenum enzymes and participate in the processes of the carbon cycle, sulfur metabolism, and nitrogen fixation in life activities (Zupok et al., 2019). In addition, intracellular molybdenum homeostasis plays an important role in the process of salt environment adaptation in *Arabidopsis thaliana* (Zupok et al., 2019; Huang et al., 2021). Thus, a molybdenum-mediated salt adaptation mechanism might also contribute to the salt acclimation process, and highlighted in Syn7942/Δsps-ect. The gene *Synpcc7942_2401* encoding the heat shock protein Hsp20 was up-regulated to 2.9 folds. It was reported that the expression of *Hsp20* gene of *Oryza sativa* had a positive response to heat and high-salt environment, and the *Hsp20* gene of rice could improve the heat tolerance and salt tolerance of *E. coli* and *Pichia pastoris* (Guo et al., 2020). Therefore, *Hsp20* (*Synpcc7942_2401*) might contribute to the salt acclimation process of Syn7942.

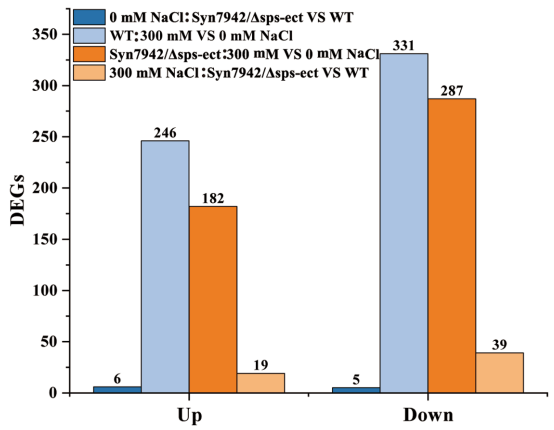

FIGURE 4
DEGs identified in transcriptomic analysis. WT: 300mM *VS* 0mM NaCl indicates the DEGs of WT under 300mM NaCl compared with 0mM NaCl condition; Syn7942/Δsps-ect: 300mM *VS* 0mM NaCl indicates the DEGs of Syn7942/Δsps-ect under 300mM NaCl compared with 0mM NaCl condition; 0mM NaCl: Syn7942/Δsps-ect *VS* WT indicates the DEGs of Syn7942/Δsps-ect grown in 0mM NaCl condition compared to WT; 300mM NaCl: Syn7942/Δsps-ect *VS* WT indicates the DEGs of Syn7942/Δsps-ect grown in 300mM NaCl condition compared to WT.

Among them, KEGG pathway enrichment analysis was then performed separately for the up-regulated genes of Syn7942/Δsps-ect. The results showed that down-regulated genes were mainly enriched in "sulfur metabolism," "energy metabolism," "ABC transporters," "membrane transport," "environmental information processing," "transporters' and "protein families: signaling and cellular processes" (Figure 6). The up-regulated genes of Syn7942/Δsps-ect was not enriched for any pathway.

Sulfur metabolism was enriched among down-regulated DEGs of Syn7942/Δsps-ect. There were 7 down-regulated genes related to sulfur

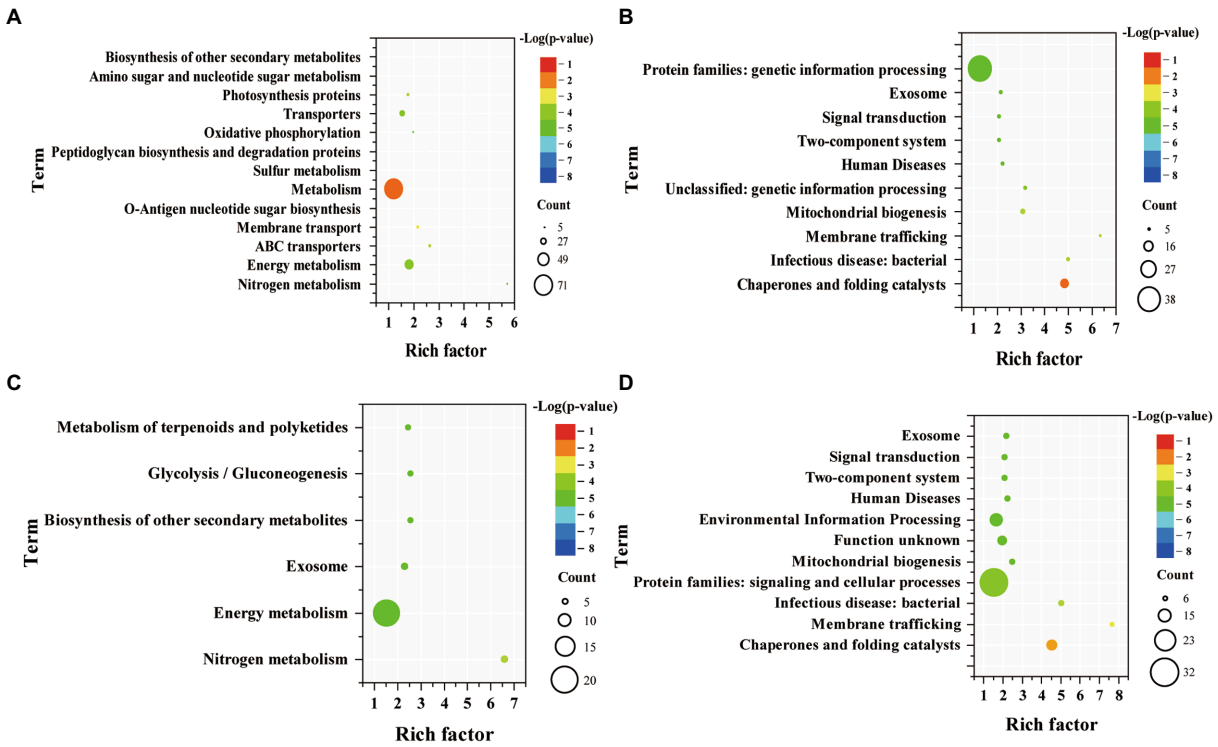

FIGURE. 5
Enriched pathway terms of DEGs in WT and Syn7942/Δsps-ect. **(A)** Enriched pathway terms based on up-regulated genes of WT; **(B)** Enriched pathway terms based on down-regulated genes of WT; **(C)** Enriched pathway terms based on up-regulated genes of Syn7942/Δsps-ect; **(D)** Enriched pathway terms based on down-regulated genes of Syn7942/Δsps-ect. WT: 300mM *VS* 0mM NaCl indicates the DEGs of WT under 300mM NaCl compared with 0mM NaCl condition; Syn7942/Δsps-ect: 300mM *VS* 0mM NaCl indicates the DEGs of Syn7942/Δsps-ect under 300mM NaCl compared with 0mM NaCl condition; 0mM NaCl: Syn7942/Δsps-ect *VS* WT indicates the DEGs of Syn7942/Δsps-ect grown in 0mM NaCl condition compared to WT; 300mM NaCl: Syn7942/Δsps-ect *VS* WT indicates the DEGs of Syn7942/Δsps-ect grown in 300mM NaCl condition compared to WT.

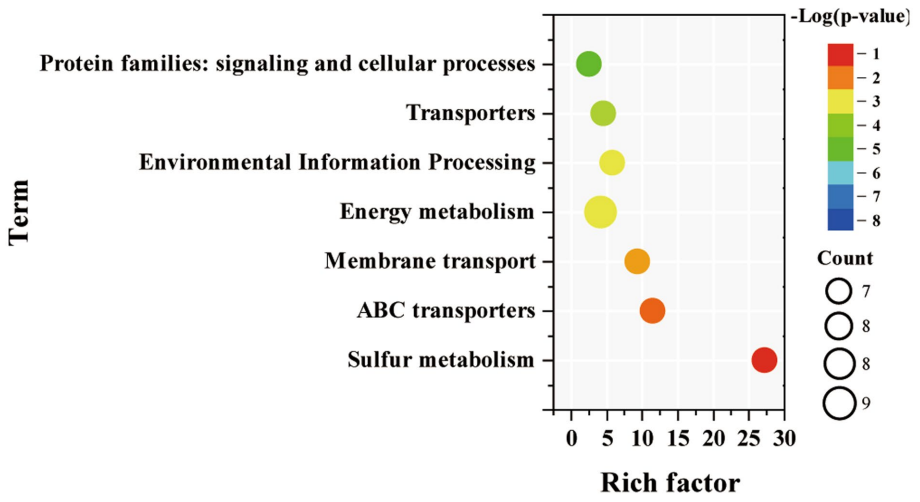

FIGURE. 6
Enriched pathway terms of down-regulated DEGs in Syn7942/Δsps-ect grown in 300mM NaCl condition compared to WT.

metabolism, namely *Synpcc7942_1681*, *Synpcc7942_1686*, *Synpcc7942_1722*, *Synpcc7942_1682*, *Synpcc7942_1687*, *Synpcc7942_1688* encoding sulfur transporters, and *Synpcc7942_1689* encoding thiosulfate/3-mercaptopyruvate sulfurtransferase. These 6 DEGs encoding sulfur transporters were also enriched in ABC transporters. This indicated that the Syn7942/Δsps-ect might reduce sulfur requirement during salt adaptation. Sulfur metabolism plays a vital role in the process of biological stress resistance. Sulfur has been reported to have essential functions in plants' salt adaptation, active oxygen elimination, and heat adaptation (Ali et al., 2021; Jahan et al., 2021; Zhou et al., 2022), while cyanobacteria have a similar sulfur metabolism as plants (Kharwar et al., 2021). Consistantly, sulfur metabolism-related genes were significantly up-regulated in WT and Syn7942/Δsps-ect compared to reference conditions (Figures 5A,C). Notably, the expression levels of genes related to sulfur metabolism were lower in Syn7942/Δsps-ect than in WT (Figure 6), indicating that ectoine synthesis might reduce the intensity of sulfur metabolism which was required for salt adaptation. This might be due to the protective effect of ectoine on protein (Bilstein et al., 2021), which reduced protein damage during salt adaptation.

In this study, to improve the salt tolerance of the model cyanobacterium Syn7942, the first study of heterologous synthesis of ectoine in cyanobacteria was achieved, and the salt tolerance of Syn7942 was successfully improved. Then, the recovery of salt tolerance of the Syn7942/Δsps-ect strain was achieved by introducing the ectoine synthesis pathway into the sucrose synthesis deficient strain (Syn7942/Δsps), indicating that ectoine was able to substitute for sucrose to combat against salt stress in Syn7942. After the replacement of sucrose by ectoine, the metabolic changes in the Syn7942/Δsps-ect strain was analyzed by transcriptomic analysis, which would help understand the salt adaptation mechanism of Syn7942/sps-ect. This study provided valuable information for understanding the salt tolerance mechanism of Syn7942, as well as potential candidates targets for further engineering the salt tolerance of Syn7942.

Data availability statement

The datasets presented in this study can be found in online repositories. The names of the repository/repositories and accession number(s) can be found below: GEO GSE222067.

Author contributions

LC, TS, and WZ: conceived and designed the study. ZD: performed the experiments. ZD, TS, WZ, and LC: analyzed the data and wrote the manuscript. All authors read and approved the manuscript.

Funding

This research was supported by grants from the National Key Research and Development Program of China (Grant nos. 2021YFA0909700, 2020YFA0906800, 2018YFA0903600 and 2019YFA0904600).

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

Supplementary material

The Supplementary material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fmicb.2023.1123081/full#supplementary-material

SUPPLEMENTARY FIGURE S1
Growth curves of WT, Syn7942/NSI-ect, Syn7942/Δsps, and Syn7942/Δsps-ect. **(A)** Growth curves of WT under 0, 100, 200, 300, 400, 500 mM NaCl; **(B)** Growth curves of Syn7942/NSI-ect under 0, 100, 200, 300, 400, 500 mM NaCl; **(C)** Growth curves of, Syn7942/Δsps under 0, 100, 200, 300, 400, 500 mM NaCl; **(D)** Growth curves of, Syn7942/-sps-ect under 0, 100, 200, 300, 400, 500 mM NaCl.

SUPPLEMENTARY FIGURE S2
Standard curve of ectoine external standard method.

SUPPLEMENTARY FIGURE S3
The number of DEGs in WTand Syn7942/Δsps-ect. **(A)** The number of up-regulated genes; **(B)** The number of down-regulated genes. WT: 300 mM VS 0 mM NaCl indicates the DEGs of WT under 300 mM NaCl compared with 0 mM NaCl condition; Syn7942/Δsps-ect: 300 mM VS 0 mM NaCl indicates the DEGs of Syn7942/Δsps-ect under 300 mM NaCl compared with 0 mM NaCl condition; 0 mM NaCl: Syn7942/Δsps-ect VS WT indicates the DEGs of Syn7942/Δsps-ect grown in 0 mM NaCl condition compared to WT; 300 mM NaCl: Syn7942/Δsps-ect VS WT indicates the DEGs of Syn7942/Δsps-ect grown in 300 mM NaCl condition compared to WT.

References

Ali, M. M., Waleed Shafique, M., Gull, S., Afzal Naveed, W., Javed, T., Yousef, A. F., et al. (2021). Alleviation of heat stress in tomato by exogenous application of sulfur. *Horticulturae* 7:21. doi: 10.3390/horticulturae7020021

Billis, K., Billini, M., Tripp, H. J., Kyrpides, N. C., and Mavromatis, K. (2014). Comparative transcriptomics between *Synechococcus* PCC 7942 and *Synechocystis* PCC 6803 provide insights into mechanisms of stress acclimation. *PLoS One* 9:e109738. doi: 10.1371/journal.pone.0109738

Bilstein, A., Heinrich, A., Rybachuk, A., and Mösges, R. (2021). Ectoine in the treatment of irritations and inflammations of the eye surface. *Biomed. Res. Int.* 2021, 8885032–8885016. doi: 10.1155/2021/8885032

Calderón, M. I., Vargas, C., Rojo, F., Iglesias-Guerra, F., Csonka, L. N., Ventosa, A., et al. (2004). Complex regulation of the synthesis of the compatible solute ectoine in the halophilic bacterium *Chromohalobacter salexigens* DSM 3043^{T}. *Microbiology* 150, 3051–3063. doi: 10.1099/mic.0.27122-0

Chen, C., Chen, H., Zhang, Y., Thomas, H. R., Frank, M. H., He, Y., et al. (2020). TBtools: an integrative toolkit developed for interactive analyses of big biological data. *Mol. Plant* 13, 1194–1202. doi: 10.1016/j.molp.2020.06.009

Chen, J., Liu, P., Chu, X., Chen, J., Zhang, H., Rowley, D. C., et al. (2020). Metabolic pathway construction and optimization of *Escherichia coli* for high-level Ectoine production. *Curr. Microbiol.* 77, 1412–1418. doi: 10.1007/s00284-020-01888-6

Cui, J., Sun, T., Chen, L., and Zhang, W. (2020). Engineering salt tolerance of photosynthetic cyanobacteria for seawater utilization. *Biotechnol. Adv.* 43:107578. doi: 10.1016/j.biotechadv.2020.107578

Cui, J., Sun, T., Chen, L., and Zhang, W. (2021). Salt-tolerant *Synechococcus elongatus* UTEX 2973 obtained via engineering of heterologous synthesis of compatible solute Glucosylglycerol. *Front. Microbiol.* 12:650217. doi: 10.3389/fmicb.2021.650217

Davies, F. K., Fricker, A. D., Robins, M. M., Dempster, T. A., McGowen, J., Charansia, M., et al. (2021). Microbiota associated with the large-scale outdoor cultivation of the cyanobacterium *Synechococcus sp.* PCC 7002. *Algal Res.* 58:102382. doi: 10.1016/j.algal.2021.102382

Farrokh, P., Sheikhpour, M., Kasaeian, A., Asadi, H., and Bavandi, R. (2019). Cyanobacteria as an eco-friendly resource for biofuel production: a critical review. *Biotechnol. Prog.* 35:e2835. doi: 10.1002/btpr.2835

Fenizia, S., Thume, K., Wirgenings, M., and Pohnert, G. (2020). Ectoine from bacterial and algal origin is a compatible solute in microalgae. *Mar. Drugs* 18:42. doi: 10.3390/md18010042

Galinski, E. A., Pfeiffer, H.-P., and Trüper, H. G. (1985). 1,4,5,6-Tetrahydro-2-methyl-4-pyrimidinecarboxylic acid. A novel cyclic amino acid from halophilic phototrophic bacteria of the genus *Ectothiorhodospira*. *Eur. J. Biochem.* 149, 135–139. doi: 10.1111/j.1432-1033.1985.tb08903.x

Giesselmann, G., Dietrich, D., Jungmann, L., Kohlstedt, M., Jeon, E. J., Yim, S. S., et al. (2019). Metabolic engineering of *Corynebacterium glutamicum* for high-level Ectoine production: design, combinatorial assembly, and implementation of a transcriptionally balanced heterologous Ectoine pathway. *Biotechnol. J.* 14:e1800417. doi: 10.1002/biot.201800417

Göller, K., Ofer, A., and Galinski, E. A. (1998). Construction and characterization of an NaCl-sensitive mutant of *Halomonas elongata* impaired in ectoine biosynthesis. *FEMS Microbiol. Lett.* 161, 293–300. doi: 10.1111/j.1574-6968.1998.tb12960.x

Gunde-Cimerman, N., Plemenitaš, A., and Oren, A. (2018). Strategies of adaptation of microorganisms of the three domains of life to high salt concentrations. *FEMS Microbiol. Rev.* 42, 353–375. doi: 10.1093/femsre/fuy009

Guo, L.-M., Li, J., He, J., Liu, H., and Zhang, H.-M. (2020). A class I cytosolic HSP20 of rice enhances heat and salt tolerance in different organisms. *Sci. Rep.* 10:1383. doi: 10.1038/s41598-020-58395-8

Huang, X.-Y., Hu, D.-W., and Zhao, F.-J. (2021). Molybdenum: more than an essential element. *J. Exp. Bot.* 73, 1766–1774. doi: 10.1093/jxb/erab534

Jahan, B., Rasheed, F., Sehar, Z., Fatma, M., Iqbal, N., Masood, A., et al. (2021). Coordinated role of nitric oxide, ethylene, nitrogen, and sulfur in plant salt stress tolerance. *Stress* 1, 181–199. doi: 10.3390/stresses1030014

Kato, Y., Inabe, K., Hidese, R., Kondo, A., and Hasunuma, T. (2022). Metabolomics-based engineering for biofuel and bio-based chemical production in microalgae and cyanobacteria: a review. *Bioresour. Technol.* 344:126196. doi: 10.1016/j.biortech.2021.126196

Kauth, M., and Trusova, O. V. (2022). Topical Ectoine application in children and adults to treat inflammatory diseases associated with an impaired skin barrier: a systematic review. *Dermatol. Ther.* 12, 295–313. doi: 10.1007/s13555-021-00676-9

Kharwar, S., Bhattacharjee, S., Chakraborty, S., and Mishra, A. K. (2021). Regulation of sulfur metabolism, homeostasis and adaptive responses to sulfur limitation in cyanobacteria. *Biologia* 76, 2811–2835. doi: 10.1007/s11756-021-00819-5

Kim, W. J., Lee, S.-M., Um, Y., Sim, S. J., and Woo, H. M. (2017). Development of SyneBrick vectors as a synthetic biology platform for gene expression in *Synechococcus elongatus* PCC 7942. *Front. Plant Sci.* 8:293. doi: 10.3389/fpls.2017.00293

Klähn, S., and Hagemann, M. (2011). Compatible solute biosynthesis in cyanobacteria. *Environ. Microbiol.* 13, 551–562. doi: 10.1111/j.1462-2920.2010.02366.x

Klähn, S., Mikkat, S., Riediger, M., Georg, J., Hess, W. R., and Hagemann, M. (2021). Integrative analysis of the salt stress response in cyanobacteria. *Biol. Direct* 16:26. doi: 10.1186/s13062-021-00316-4

Li, S., Sun, T., Xu, C., Chen, L., and Zhang, W. (2018). Development and optimization of genetic toolboxes for a fast-growing cyanobacterium *Synechococcus elongatus* UTEX 2973. *Metab. Eng.* 48, 163–174. doi: 10.1016/j.ymben.2018.06.002

Liu, M., Liu, H., Shi, M., Jiang, M., Li, L., and Zheng, Y. (2021). Microbial production of ectoine and hydroxyectoine as high-value chemicals. *Microb. Cell Factories* 20:76. doi: 10.1186/s12934-021-01567-6

Liu, X., Xie, H., Roussou, S., and Lindblad, P. (2022). Current advances in engineering cyanobacteria and their applications for photosynthetic butanol production. *Curr. Opin. Biotechnol.* 73, 143–150. doi: 10.1016/j.copbio.2021.07.014

Ning, Y., Wu, X., Zhang, C., Xu, Q., Chen, N., and Xie, X. (2016). Pathway construction and metabolic engineering for fermentative production of ectoine in *Escherichia coli*. *Metab. Eng.* 36, 10–18. doi: 10.1016/j.ymben.2016.02.013

Schwibbert, K., Marin-Sanguino, A., Bagyan, I., Heidrich, G., Lentzen, G., Seitz, H., et al. (2011). A blueprint of ectoine metabolism from the genome of the industrial producer *Halomonas elongata* DSM 2581T. *Environ. Microbiol.* 13, 1973–1994. doi: 10.1111/j.1462-2920.2010.02336.x

Sengupta, A., Sunder, A. V., Sohoni, S. V., and Wangikar, P. P. (2019). Fine-tuning native promoters of *Synechococcus elongatus* PCC 7942 to develop a synthetic toolbox for heterologous protein expression. *ACS Synth. Biol.* 8, 1219–1223. doi: 10.1021/acssynbio.9b00066

Singh, M., Sharma, N. K., Prasad, S. B., Yadav, S. S., Narayan, G., and Rai, A. K. (2013). The freshwater cyanobacterium *Anabaena doliolum* transformed with *ApGSMT-DMT* exhibited enhanced salt tolerance and protection to nitrogenase activity, but became halophilic. *Microbiology* 159, 641–648. doi: 10.1099/mic.0.065078-0

Song, K., Tan, X., Liang, Y., and Lu, X. (2016). The potential of *Synechococcus elongatus* UTEX 2973 for sugar feedstock production. *Appl. Microbiol. Biotechnol.* 100, 7865–7875. doi: 10.1007/s00253-016-7510-z

Soontharapirakkul, K., Promden, W., Yamada, N., Kageyama, H., Incharoensakdi, A., Iwamoto-Kihara, A., et al. (2011). Halotolerant cyanobacterium Aphanothece halophytica contains an Na^{+}-dependent F1F0-ATP synthase with a potential role in salt-stress tolerance. *J. Biol. Chem.* 286, 10169–10176. doi: 10.1074/jbc.M110.208892

Sun, T., Li, S., Song, X., Diao, J., Chen, L., and Zhang, W. (2018). Toolboxes for cyanobacteria: recent advances and future direction. *Biotechnol. Adv.* 36, 1293–1307. doi: 10.1016/j.biotechadv.2018.04.007

Tan, C., Xu, P., and Tao, F. (2022). Carbon-negative synthetic biology: challenges and emerging trends of cyanobacterial technology. *Trends Biotechnol.* 40, 1488–1502. doi: 10.1016/j.tibtech.2022.09.012

Waditee-Sirisattha, R., Singh, M., Kageyama, H., Sittipol, D., Rai, A. K., and Takabe, T. (2012). *Anabaena sp.* PCC7120 transformed with glycine methylation genes from Aphanothece halophytica synthesized glycine betaine showing increased tolerance to salt. *Arch. Microbiol.* 194, 909–914. doi: 10.1007/s00203-012-0824-z

Wang, X., Jin, G., Pan, K., Zhu, B., and Li, Y. (2021). Effects of fluctuating temperature in open raceway ponds on the biomass accumulation and harvest efficiency of spirulina in large-scale cultivation. *Environ. Sci. Pollut. Res.* 28, 20794–20802. doi: 10.1007/s11356-020-11914-6

Yu, J., Wang, Z., Wang, J., Mohisn, A., Liu, H., Zhang, Y., et al. (2022). Physiological metabolic topology analysis of *Halomonas elongata* DSM 2581^{T} in response to sodium chloride stress. *Biotechnol. Bioeng.* 119, 3509–3525. doi: 10.1002/bit.28222

Zhang, H., Liang, Z., Zhao, M., Ma, Y., Luo, Z., Li, S., et al. (2022). Metabolic engineering of *Escherichia coli* for Ectoine production with a fermentation strategy of supplementing the amino donor. *Front. Bioeng. Biotechnol.* 10:824859. doi: 10.3389/fbioe.2022.824859

Zhang, M., Luo, Q., Sun, H., Fritze, J., Luan, G., and Lu, X. (2022). Engineering a controllable targeted protein degradation system and a derived OR-GATE-type inducible gene expression system in *Synechococcus elongatus* PCC 7942. *ACS Synth. Biol.* 11, 125–134. doi: 10.1021/acssynbio.1c00226

Zhou, X., Joshi, S., Patil, S., Khare, T., and Kumar, V. (2022). Reactive oxygen, nitrogen, carbonyl and sulfur species and their roles in plant abiotic stress responses and tolerance. *J. Plant Growth Regul.* 41, 119–142. doi: 10.1007/s00344-020-10294-y

Zupok, A., Iobbi-Nivol, C., Méjean, V., and Leimkühler, S. (2019). The regulation of Moco biosynthesis and molybdoenzyme gene expression by molybdenum and iron in bacteria. *Metallomics* 11, 1602–1624. doi: 10.1039/c9mt00186g

frontiers | Frontiers in Microbiology

TYPE Review
PUBLISHED 03 February 2023
DOI 10.3389/fmicb.2023.1111794

OPEN ACCESS

EDITED BY
Martin Hagemann,
University of Rostock,
Germany

REVIEWED BY
Jianhua Fan,
East China University of Science and Technology, China
Yandu Lu,
Hainan University,
China

*CORRESPONDENCE
Qiang Wang
✉ wangqiang@henu.edu.cn

SPECIALTY SECTION
This article was submitted to Microbiotechnology, a section of the journal Frontiers in Microbiology

RECEIVED 30 November 2022
ACCEPTED 06 January 2023
PUBLISHED 03 February 2023

CITATION
Hu X, Fan Y, Mao C, Chen H and Wang Q (2023) Application of transposon insertion site sequencing method in the exploration of gene function in microalgae.
Front. Microbiol. 14:1111794.
doi: 10.3389/fmicb.2023.1111794

Application of transposon insertion site sequencing method in the exploration of gene function in microalgae

Xiaobing Hu[1,2], Yulong Fan[1], Chengfeng Mao[1], Hui Chen[1] and Qiang Wang[1,3]*

[1]State Key Laboratory of Crop Stress Adaptation and Improvement, School of Life Sciences, Henan University, Kaifeng, China, [2]School of Environmental Engineering, Yellow River Conservancy Technical Institute, Kaifeng, China, [3]Academy for Advanced Interdisciplinary Studies, Henan University, Kaifeng, China

Microalgae are a large group of organisms that can produce various useful substances through photosynthesis. Microalgae need to be genetically modified at the molecular level to become "Chassis Cells" for food, medicine, energy, and environmental protection and, consequently, obtain benefits from microalgae resources. Insertional mutagenesis of microalgae using transposons is a practical possibility for understanding the function of microalgae genes. Theoretical and technical support is provided in this manuscript for applying transposons to microalgae gene function by summarizing the sequencing method of transposon insertion sites.

KEYWORDS

microalgae, chassis cell, transposon, flanking sequence, sequencing

1. Introduction

Microalgae are one of the oldest groups of organisms on Earth, contributing more than 50% of the primary productivity of the entire planet (Sun et al., 2022). Compared with other biomass resources, microalgae occupy cultivated lands, have high biomass, grow at a fast rate, have high adaptability, are easy to domesticate, and have high light energy utilization. Additionally, through genetic transformation, engineered microalgal chassis cells can fix CO_2 (Singh and Ahluwalia, 2012) through photosynthesis to produce substances, including oils, proteins, amino acids, polysaccharides, and vitamins. Currently, microalgae have been widely used in food (Scieszka and Klewicka, 2019; Fu et al., 2021), medicine (Beaumont et al., 2021), energy (Zhang et al., 2012; Frigon et al., 2013; Bigelow et al., 2014), environmental protection (Cabanelas et al., 2013; Gomez et al., 2013; Chen W. et al., 2016), feed (Vidyashankar et al., 2014; Packer et al., 2016), and other fields. Therefore, these organisms gradually became critical raw materials for the active extraction of substances (Liu et al., 2022). Engineering microalgal chassis cells will become an effective force in achieving the goal of carbon neutralization worldwide and, consequently, replacing traditional industries.

It is essential to further understand the gene functions of microalgae in depth to utilize microalgae resources. However, the gene functions of a considerable proportion of microalgae remain unknown. More methods are being used to analyze and identify gene functions with the continuous development and innovation of new molecular biology technologies and methods (Ng et al., 2020). The functional genomics sub-discipline gradually formed after such approaches were developed. The construction of effective mutants is an essential method in functional genomics research. There are many methods for obtaining mutants of genes, and transposons to construct mutants have a random nature. This method may better understand gene functions and the connections between related genes (van Opijnen and Camilli, 2013).

Transposons are mobile DNA genetic sequences that can "jump" to distinct locations in the genome and are found in prokaryotic and eukaryotic genomes (Choi and Kim, 2009). Barbara McClintock discovered the first transposon in maize (Gierl and Saedler, 1992). Transposon tags have long been considered a powerful research tool for randomly distributing primer binding sites, generating mutations, and introducing physical or genetic tags into large target DNA (Damasceno et al., 2010; Shapiro, 2010). Therefore, the random insertion of transposon mutations is a desirable choice, especially if one wants to create many mutants.

After the insertion of a mutation is completed in the transposon, the gene identification and location of the insertion mutation must be solved (Li et al., 2020). This means understanding the gene sequences on either side of the insertion site of the mutant by sequencing. We can only understand the function that a gene may have through the correlation between the mutation position and the phenotype. Therefore, standard sequencing methods involving microalgae are introduced and summarized in this study.

2. Enzymatic digestion

Restriction endonucleases are used in the enzymatic digestion method to digest the microalgae genome before amplification and sequencing. This method is straightforward, has low costs, and is easy to operate, but the success rate is low. It is suitable for mutants with a negligible overall genome and appropriate restriction endonucleases. Standard methods are described as follows.

2.1. Reverse PCR

Reverse PCR is used to find a restriction endonuclease with more enzymatic sites and broader distribution in the mutant genome, but no enzymatic sites or only one enzymatic site in the transposon sequence or fragment the genome by enzymatic digestion. The DNA fragment is self-associated after enzymatic digestion by ligase to cyclize it. Specific primers can be designed from the transposons if the cyclized genome contains transposons. Specific primers can be designed from the transposon for amplification and sequencing to obtain the transposon insertion site if the transposon is included in the genome. The sequence obtained is on both sides of the insertion site if there is no enzyme cut site in the transposon (Figure 1A). Finally, the sequence obtained is on one side of the insertion site if there is a single enzyme-cut site (Figure 1B).

The reverse PCR method is simple in principle and operation, and its experimental cost is low. However, it is unsuitable for high-throughput sequencing. Additionally, it is not stable as a sequencing method because of its specific requirements for selecting restricted endonucleases and restrictions on the transposon insertion position. In addition, its sequencing length is not fixed. Feng et al. (2008) used reverse PCR to amplify the 5′ and 3′ flanking sequences of the TaCKX1 gene. These researchers obtained the full-length DNA sequence of TaCKX1 by cloning the TaCKX1 fragment from a conserved sequence of wheat cytokinin oxidase/dehydrogenase (CKX). Wang et al. (2021) used reverse PCR to amplify and detect the FSTA gene in transgenic zebrafish genomic DNA. This method may be used for gene doping detection in blood samples or to assess the safety of gene therapy and GMOs. Su et al. (2021) used this assay for transposase-accessible chromatin using the sequencing (ATAC-seq) method combined with the "Circle_finder" bioinformatic algorithm to predict extrachromosomal circular DNA (eccDNA) in human cancer cells. These researchers validated the detection of eccDNA using reverse PCR. Gutiérrez et al. (2021) used reverse PCR to identify chronic myeloid leukemia (CML) at the genomic level with the breakpoint sequence of the signature fusion gene *BCR-ABL1*. They applied this method to seven real cases.

2.2. Plasmid rescue

The principle of plasmid rescue is similar to reverse PCR. The standard operation is (1) to insert the transposon into the genome, (2) digest it with a restriction endonuclease, (3) ligate it into a cloning vector, (4) transform it into an *Escherichia coli* culture, (5) select positive bacteria according to the label carried by the plasmid, (6) culture it, (7) extract its plasmid, and (8) sequence it according to the specific primers of the transposon and plasmid (Tsurumaru et al., 2008; Figure 2).

The advantage of the plasmid rescue method is the high specificity of the fragments obtained. The disadvantages of this method are the requirement for the selection of endonucleases, the lack of experimental stability, and the unsuitability for large-scale high-throughput sequencing. Huang et al. (2009) used a combination of plasmid rescue and reverse PCR to simultaneously determine the sequences on both sides of the *Drosophila* P-transposon insertion site. Kemppainen et al. (2008) randomly determined the genomic DNA sequences of the T-DNA right border (Rb) of 51 strains from a considerable number (~500) of T-DNA insertion mutants of *Laccaria bicolor* using plasmid rescue. Sixty-nine percent of the flanking sequences of this species were successfully determined. At the same time, 87% of these sequences were successfully localized in the genome.

2.3. Specific enzymatic cleavage

Some specific restriction endonucleases are used in this method. These endonucleases have a common point: the enzymatic cut site is located after several bases of the recognition site. Therefore, a segment of the base sequence of the recognition site can be left after the enzymatic cut. For example, MmeI (Figure 3A) and EcoP15I (Figure 3B) can retain about 18 ~ 27 bp after the recognition site after enzymatic cleavage. Therefore, these enzymatic cleavage sites can be inserted at both ends of the transposon. The insertion site was sequenced by amplifying the specific sequences on the specific motifs by enzymatic ligation or by directly ligating special connectors. Finally, the insertion site was found against the target genome (Figure 3C).

The advantage of this method is that it is simple. Only particular enzyme cleavage sites must be added on both sides of the transposon. It can support high-throughput large-scale sequencing by connecting Illumina adapter sequences. The disadvantage of this method is related to the short localization of the sequence, which is only about 38~52 bp. Therefore, sometimes even the genome length that can be measured is less than 38 bp to consider joint connection and sequencing problems. The specificity of the method is weak, and the position of the transposon insertion mutation cannot be accurately determined in some genomes with more repetitive sequences or palindromic sequences. Additionally, the particular enzyme cleavage method cannot be used for transposons, in which transposase

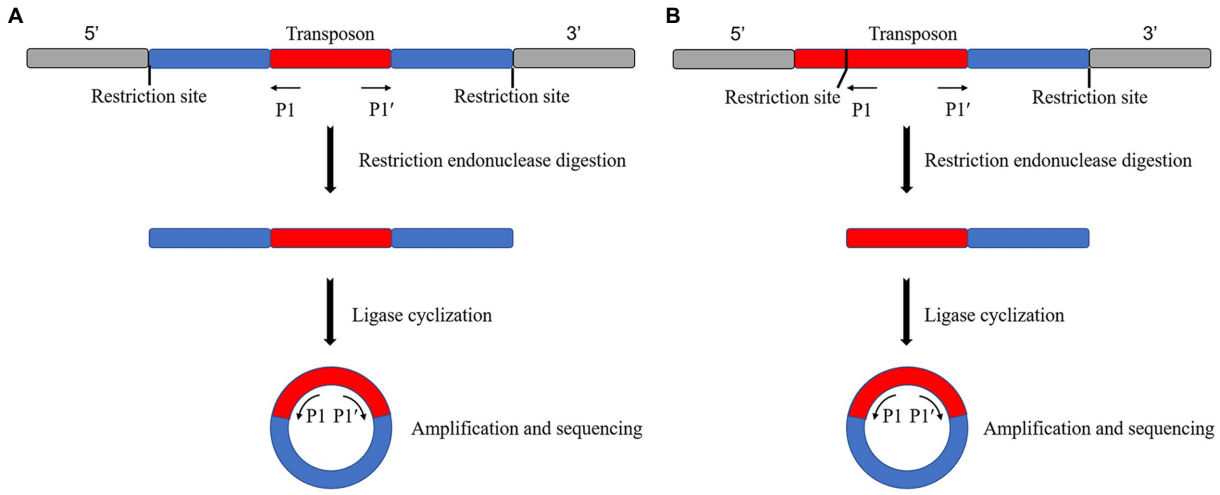

FIGURE 1
Schematic diagram of the reverse PCR principle. **(A)** No restriction endonuclease digestion site on the transposon, sequenced as both sides of the transposon insertion site after self-associative cyclization. **(B)** Single restriction endonuclease digestion site on transposon, sequenced as a unilateral sequence of the transposon insertion site after self-associative cyclization.

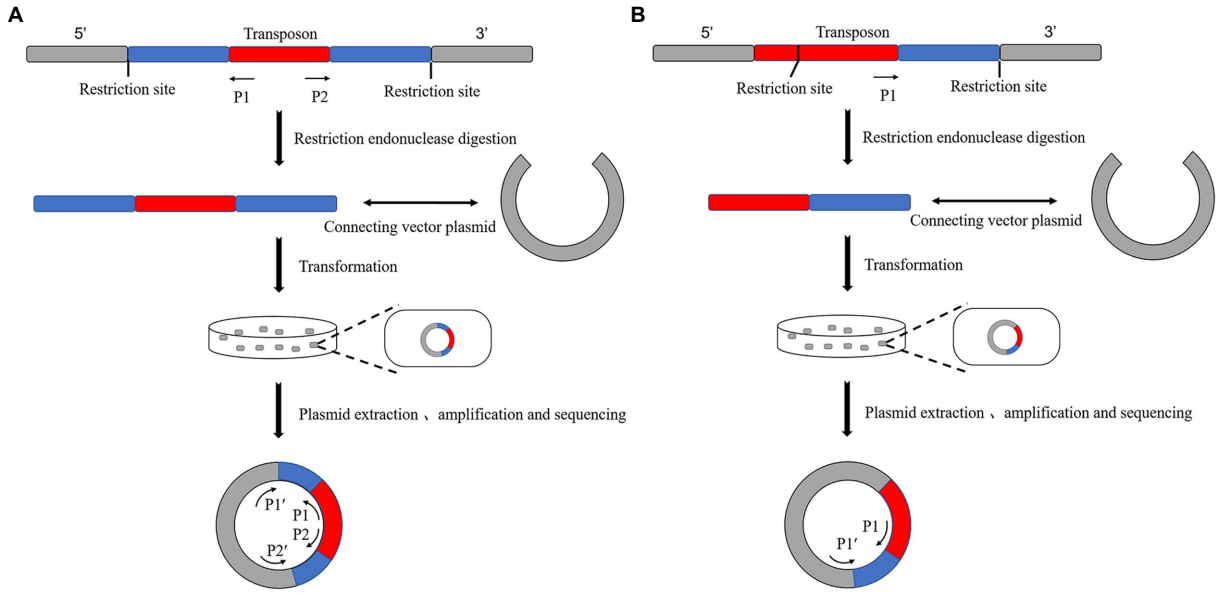

FIGURE 2
Schematic diagram of the principle of plasmid rescue. **(A)** There is no restriction endonuclease site on the transposon. The vector plasmid was ligated after enzymatic digestion and transformed into *E. coli*. The positive new bacteria were obtained by selective labeling. The plasmid is extracted, and the sequence of both sides of the transposon insertion site can be measured by amplifying and sequencing two pairs of specific primers on the transposon and plasmid. **(B)** There is a single restriction endonuclease site on the transposon. The plasmid was extracted from a positive bacterium obtained using selective labeling. The unilateral sequence of the transposon insertion site could be sequenced by amplifying and sequencing the transposon with a pair of specific primers on the transposon and plasmid.

recognition sites are at both ends because the enzyme cleavage sites cannot be added.

Regarding the application of the method, Ng et al. (2005) investigated a more accurate and efficient way of determining cDNA using the MmeI endonuclease in conjunction with other common endonucleases. These researchers mapped the cDNA to genomic sequences to delineate the transcriptional boundaries of each gene. Matsumura et al. (2003) used the EcoP15I endonuclease to analyze the sequence of cDNA applied to monitor the genome sequences of rice and *P. aeruginosa*. These researchers found that hydrophobic protein genes were the most actively transcribed in *P. aeruginosa* leaves. They also studied gene expression changes in *Benthamiana*, a model organism, before the hypersensitive response induced by INF1, allowing the rapid identification of genes that were up- or down-regulated by the induction. Zhang et al. (2014) investigated a method to determine the transcriptional boundaries of a

10.3389/fmicb.2023.1111794

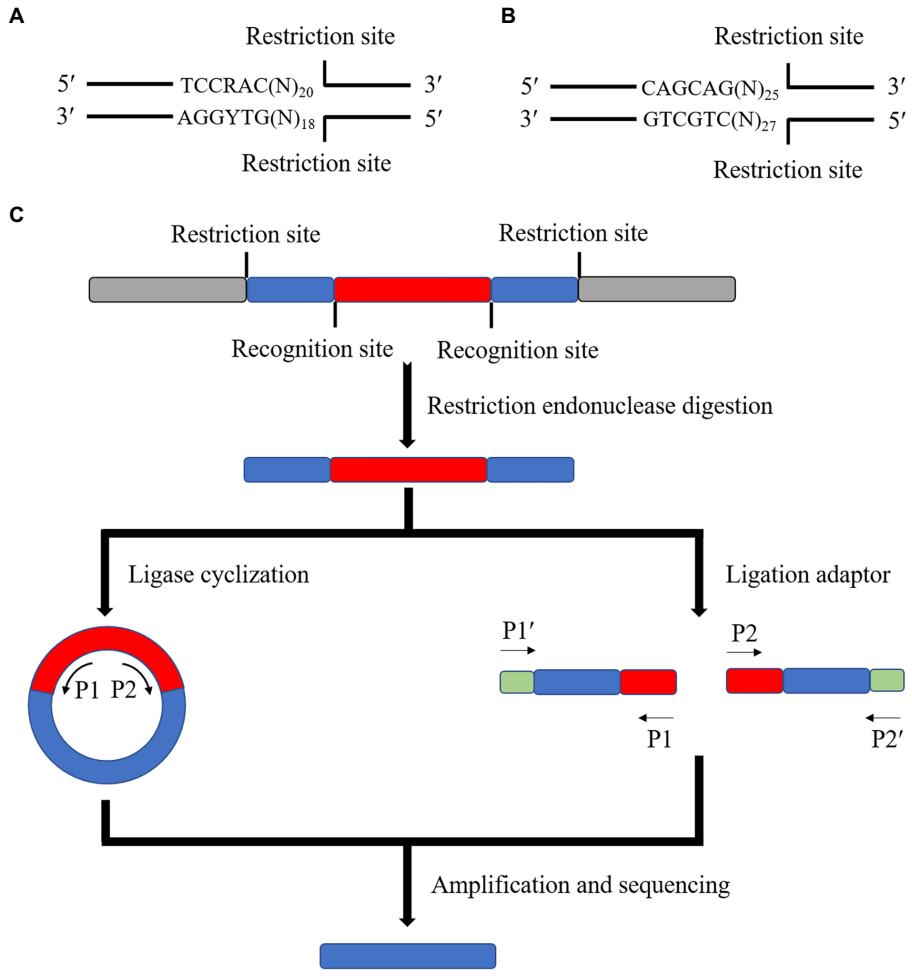

FIGURE 3
Schematic diagram of the principle of the special enzyme method. **(A)** Schematic diagram of the MmeI enzyme cleavage site (R is any purine, Y is any pyrimidine, N is any base). **(B)** Schematic diagram of the EcoP15I enzyme cleavage site. **(C)** The special enzyme cleavage method is to modify both ends of the transposon in advance and load the recognition site of a particular enzyme before transposition. After inserting the transposon into the target genome, the sequence of 18~27bp is left on both sides of the fragment containing the transposon using a special enzyme cleavage. The ligase is cyclized or connected to the sequencing junction to amplify and sequence with the specific primer or junction sequence on the transposon.

high-throughput sequencing method for determining transposon insertion sites in *Chlamydomonas reinhardtii*. The species was investigated and applied to a mutant library, and 11,478 insertion sites were identified.

3. Multiple primer amplification method

The multiple primer amplification methods are developed based on chromosome stepping and nested PCR principles. Nested PCR is a multiple primer PCR method designed to enhance the specificity of the pairing between primers and templates based on standard PCR. The principle is straightforward. The most basic nested PCR is to set two sets of PCR primers for two rounds of PCR amplification using the first pair of primers (also known as external primers) for multiple cycles of standard amplification of the target DNA. Part of the amplified product is diluted after the first amplification round and used as a template for the second round of amplification, using the second pair of primers (known as internal primers or nested primers, combined with the first round of PCR products). The second primer pair, called internal primers or nested primers, which are combined inside the PCR product of the first round, is used for multiple amplification cycles. Sometimes, a third or fourth primer pair can be used for amplification, depending on the experiment.

However, when designing primers, specific primers in nested PCR are not designed based on randomly inserted transposons. Two PCR rounds can only be completed after some universal primers are created, which often cannot be used directly in the practical application of transposon insertion mutant sequencing. Therefore, some improved methods have been derived and are described below.

3.1. Thermal asymmetric interleaving PCR (TAIL PCR)

The basic principle of TAIL PCR is the same as nested PCR. TAIL PCR is based on designing multiple sets of nested specific primers with a higher annealing temperature (Tm) on the transposon and a shorter and lower Tm value of random simpler primers. The target

10.3389/fmicb.2023.1111794

TABLE 1 TAIL PCR amplification procedures.

Reaction	Procedures		Products
	Number of cycles	Reaction conditions	
First PCR cycle	1	92°C (180 s), 95°C (60 s)	The target sequences amplify linearly by annealing and extending specific primer one with transposon sequences at high annealing temperatures. At the same time, the concentration of non-specific products resulting from the binding of the simplex primers is low.
	5	94°C (30 s), 65°C (60 s), 72°C (120 s)	
	1	94°C (30 s), 25°C (120 s), not less than 120 s to 72°C, 72°C (120 s)	The low annealing temperature allows better binding of the simplex primers to a larger number of target sequences.
	10	94°C (30 s), 44°C (60 s), 72°C (120 s)	The lower annealing temperature allows both primers to anneal to the template. Thereby, it enables the original single-stranded target DNA produced by the high specificity cycle to be replicated into double-stranded DNA in preparation for the next round of amplification.
	12	94°C (30 s), 65°C (60 s), 72°C (120 s), 94°C (30 s), 65°C (60 s), 72°C (120 s), 94°C (30 s), 44°C (60 s), 72°C (120 s)	Alternating cycles of higher and lower annealing temperatures allow exponential amplification of the target fragment. This exceeds the amount of non-target fragments.
	1	72°C (300 s)	
Second PCR cycle	12	94°C (30 s), 65°C (60 s), 72°C (120 s), 94°C (30 s), 65°C (60 s), 72°C (120 s), 94°C (30 s), 45°C (60 s), 72°C (120 s)	The first round of product dilution is used as a template. Then, the specific primer two and the combination of the simplex primers allow the specific product to be selectively amplified.
	1	72°C (300 s)	
Third PCR cycle	20	94°C (40 s), 45°C (60 s), 72°C (120 s)	After the second round of product dilution as the template, the specific primer three was combined with the simplex primer. The target fragment was further specifically amplified to obtain the target sequence flanking the transposon.
	1	72°C (300 s)	

sequence was amplified by amplifying different specific primers and simpler primers using the difference in Tm values (Liu et al., 1995; Liu and Wittier, 1995).

The commonly used TAIL PCR amplification method consists of three cycles of PCR reactions (Table 1; Liu and Huang, 1998). In the first cycle, products amplified by a specific primer one and simplex primer are yielded by the PCR reaction (type I). The products were amplified by a particular primer (type II). The products are amplified by a simplex primer (type III). During the second PCR reaction cycle, the product of the first cycle was diluted as a template. The product of the type I primer is selectively amplified using specific primer two. This primer was made using a simplex primer in a thermally asymmetric supercycle. The product of the second PCR reaction cycle was diluted in the third PCR reaction cycle. A template and a specific primer from the third cycle with a simplex primer are used in a normal PCR reaction cycle or a thermally asymmetric supercycle. The target fragment was further amplified to obtain the sequence on one side of the transposon insertion site (Figure 4).

The main advantage of the TAIL PCR method is that it does not require DNA manipulation before PCR. Cyclization and ligation are avoided and have a faster reaction speed, higher specificity, and higher efficiency. However, nonspecific binding due to low temperature can still exist and may sometimes lead to amplification and sequencing failures or situations where the amplified sequence length is insufficient. Various improved versions of amplification protocols are constantly updated as TAIL PCR continues to develop. For instance, the success rate and amplification sequence length are improved by the method by setting multiple sets of more extended simplex primers, increasing the success rate to 90% and the amplification sequence length to 1–3 kb (Liu and Chen, 2007).

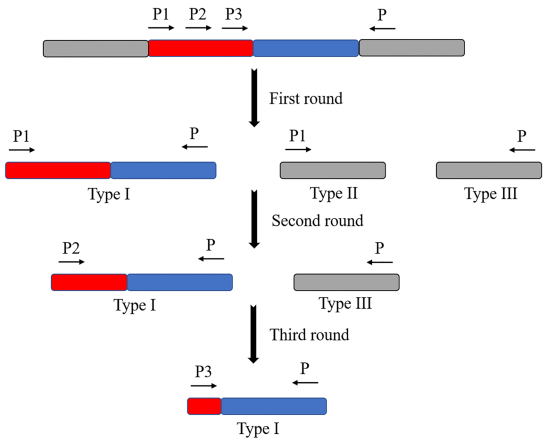

FIGURE 4
Schematic diagram of the TAIL-PCR principle. Three specific primers were designed on the transposon. One randomly abridged primer is designed to produce three products by the first amplification cycle (high specificity reaction, low specificity reaction, low specificity reaction, and thermally asymmetric super response). These are the target fragment (type I product), the fragment amplified by the specific primer itself (type II product), and the fragment amplified by the abridged primer itself (type III product). The target fragment was selectively amplified by the second amplification cycle (thermally asymmetric super reaction). The target fragment was further amplified in the third amplification cycle (thermally asymmetric super reaction or normal PCR reaction).

Yuan et al. (2009) used a combination of TAIL PCR and plasmid rescue to efficiently identify eight insertional mutation sites in a library of rice streak transposon Tn5 insertional mutants with attenuated

virulence on rice. These researchers used this method to determine the corresponding functional genes efficiently. Oranab et al. (2021) used the TAIL PCR technique to examine the T-DNA insertion sites of the activation marker mutants of the CNGC19 and CNGC20 genes in *Arabidopsis* cyclic nucleotide-gated ion channels (CNGCs) under salt stress conditions. Thus, it lays the groundwork for studying the role of CNGC19 and CNGC20 in *Arabidopsis* under salt stress regulation. Wang et al. (2013) used a modified TAIL PCR technique to examine the genome of *Wolbachia*. The WO genome of the mild phage on *Wolbachia* was determined using a modified TAIL PCR technique. The evolution of the WO genome was also assessed by comparing the WO genomes of infested fig wasps with those of infected insects. The following species were considered: the pink spotted borer moth, *Culex* mosquito, *Drosophila melanogaster*, *Drosophila anthropomorphis*, and the lyre fly nymphal set of golden wasps.

3.2. Rapid amplification of cDNA ends (RACE)

RACE is a technique based on reverse transcription PCR to rapidly amplify the 5′ and 3′ ends of cDNA from samples (Chutia et al., 2020). Since cDNA differs in prokaryotic and eukaryotic algae, and the situation is different at the 5′ and 3′ ends, various amplification methods are described below.

Reverse transcription primers were designed for eukaryotic microalgae to reverse the transcription of the first cDNA strand based on the naturally occurring poly(A) tail at the 3′ end of mRNA (Passmore and Coller, 2021). Specific primers were designed to synthesize the second cDNA strand based on transposon sequences. Subsequently, PCR amplification of the obtained cDNA strand was performed with the specific primer and the 3′ end primer of the righteous strand as a pair of primers to obtain the 3′ end sequence of cDNA (Figure 5A). In contrast, it is necessary to design specific primers based on transposon sequences, since there is no naturally recognizable sequence at the 5′ end of eukaryotic microalgae mRNA. Therefore, it will be possible to reverse transcribe it to obtain the first cDNA strand. At the same time, primer sequences at the 3′ end of cDNA by enzymatic linkage will be added, often with a poly(C) tail, and specific primers will be designed to synthesize the second cDNA strand based on the added sequence. The second cDNA strand was used as a template to synthesize double-stranded cDNA using transposon-specific primers. Finally, the cDNA 5′ end sequence was obtained by PCR amplification using transposon-specific primers and antisense strand 3′ end primers (Figure 5B).

The 3′ end of mRNA does not have a special structure similar to the poly(A) tail for prokaryotic microalgae. Therefore, a splice sequence must be directly attached to the 3′ end of the mRNA to replace the poly(A) tail. The other operations are consistent with the eukaryotic microalgae 3′ end in RACE (Figure 6A). The 5′ end of prokaryotic microalgae in RACE is the same as that of eukaryotic microalgae. This requires the addition of a splice sequence at the 5′ end of the cDNA after reverse transcription and amplification (Figure 6B).

Meslet-Cladière and Vallon (2012) determined the flanking sequences of 38 randomly selected insertion mutants in a transposon insertion mutation library of the model organism *C. reinhardtii*. These authors used the 3′ end of the RACE technique. Twenty-seven (71%) were valid flanking sequences, and 23 could be accurately localized in the genome. Hu et al. (2017) identified small regulatory RNAs (SRNAs) in *Synechocystis* sp. PCC 6803 uses 5′ and 3′ ends in the RACE method, naming it RblR. RblR positively regulates the gene *rbcL*. *rbcL* encodes a large chain of Rubisco, an enzyme that catalyzes carbon fixation under different stress conditions. Thus, it affects photosynthesis regulation in PCC 6803. Li et al. (2022) determined the sequence of small antisense RNA (ThfR) on the reverse complementary strand of the *sll1414* (*thf1*) gene in PCC 6803 was used in the 5′ and 3′ ends in the RACE technique. These researchers investigated the relationship between ThfR and gene *thf1* by examining its high- and low-expression mutants.

3.3. Linear amplification-mediated PCR (LAM-PCR)

Target products are obtained in the linear amplification mediated-PCR (LAM-PCR) method by designing multiple primer sets amplified step by step. The first step is to amplify single-stranded DNA using transposon-specific primers with biotin. The amplified single-stranded DNA is captured by the adsorption of biotin by streptavidin magnetic beads. The insertion site flanking sequence was obtained by amplification and sequencing (Figure 7). This method is precise, has a high success rate, and may be designed to link Illumina junctions in the second round of amplification primers if needed (Carette et al., 2011). However, this method is more expensive for sequencing individual mutants, if not high-throughput sequencing.

Schmidt et al. (2007) used LAM-PCR to detect integration sites representing unique molecular markers for each transduced cell and its clonal progeny in the cells of an integration vector system for clinical gene therapy. Gabriel et al. (2014) used LAM-PCR to demonstrate that leukemia originated from the provirus-induced overexpression of adjacent proto-oncogenes in gene therapy patients. It was possible to bypass restriction digestion with LAM-PCR, eliminating retrieval bias at the integration site. This enabled a comprehensive analysis of the provirus location in the host genome, detailing a stepwise amplification method that integrates adjacent 3′ and 5′ sequences of the lentiviral vector.

4. Transposon mutagenesis coupled with next-generation sequencing (Tn-Seq)

Transposon mutation combined with next-generation sequencing (Tn-Seq) is a high-throughput analysis method for transposon insertion. The basic idea of this method is to (1) physically or enzymatically interrupt the genome of the inserted transposon, (2) ligate the splice sequence required for next-generation sequencing to each fragment, and (3) amplify the specific sequence on one side of the transposon and the splice sequence on the corresponding side as primers. These steps were made to obtain DNA fragments of an appropriate size and perform next-generation sequencing (Figure 8). There are many other conceptually similar methods, including Tradis, HITS, INSeq, and TnLE-Seq (Wetmore et al., 2015). These methods have a common feature in that many transposon mutants are mixed. The abundance of transposon insertion into each gene may only be determined by high-throughput sequencing under certain growth conditions, such as the fitness of each gene under

10.3389/fmicb.2023.1111794

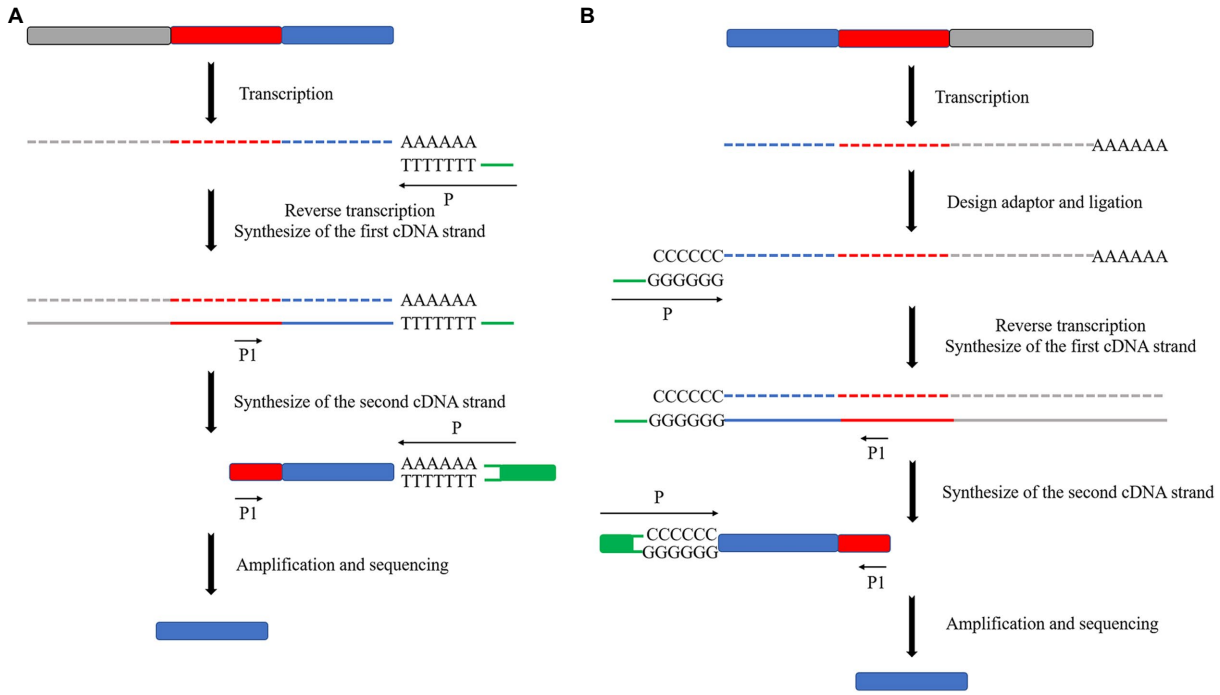

FIGURE 5
Schematic diagram of the principle of RACE in eukaryotes. **(A)** 3′ end RACE utilizes the post-transcriptional poly(A) tail structure of mRNA first to reverse transcribe the first cDNA strand containing the transposon sequence and then synthesize the second cDNA strand by using specific primers on the transposon sequence. **(B)** 5′ end RACE is performed by ligating a poly(C) tail structure after transcription. Then, the same operation as 3′ end RACE is performed.

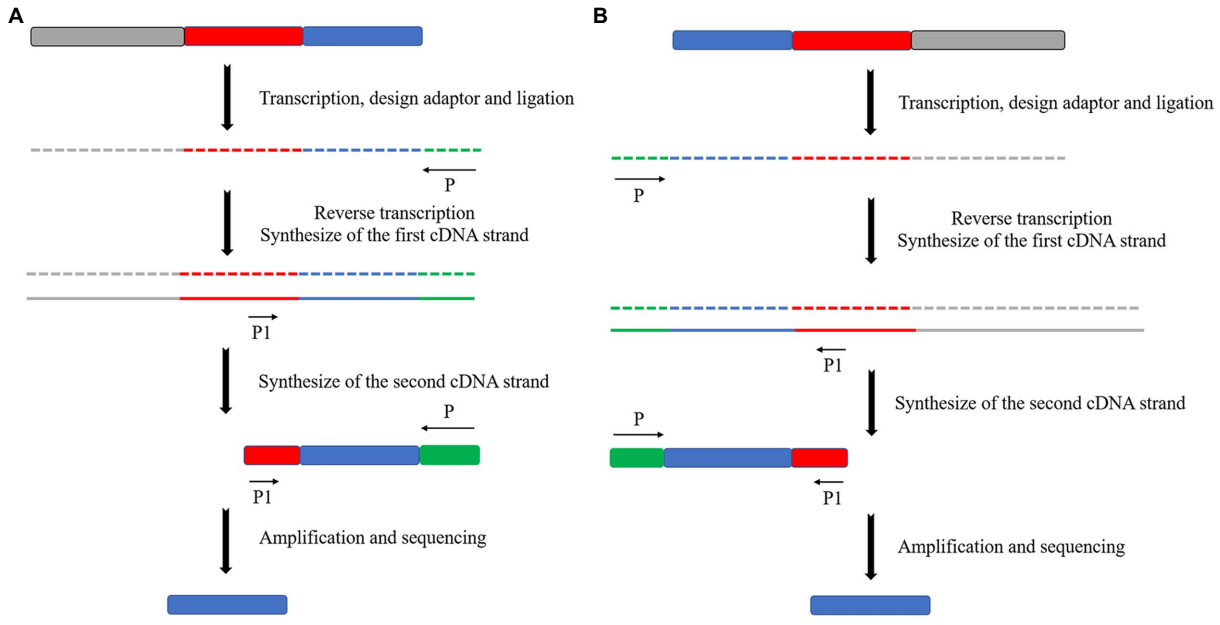

FIGURE 6
Schematic diagram of the principle of RACE in prokaryotes. **(A)** 3′ end RACE. **(B)** 5′ end RACE.

that growth condition, but by trying to separate the individual. However, it is difficult to isolate each mutant and match the insertion sites one by one.

Rubin et al. (2015) used transposons with molecular barcode "tags" into the genome of the prokaryotic microalga PCC 7942. They sequenced random molecular barcode transposon insertion mutation sites (RB-TnSeq) to create a library containing more than 250,000 transposon mutants and sequenced them to identify insertion sites. A total of 718 genes out of 2,723 were identified as necessary for the survival of the organism under laboratory conditions through an

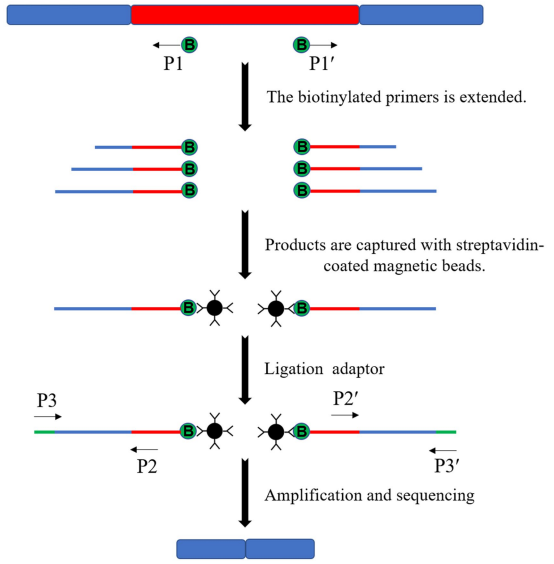

FIGURE 7
Schematic diagram of the LAM-PCR principle. Multiple single-stranded DNAs of varying lengths were amplified by the biotinylated primers. Then, multiple single-stranded DNAs are captured by streptavidin-coated magnetic beads to ligate the splice sequences and complete amplification and sequencing.

analysis of the distribution and survival of these mutants. Li et al. (2019) generated a mutant library of eukaryotic microalgae *C. reinhardtii* by adding a DNA barcode to transposons 3′ and 5′ respectively through RB TnSeq. The library has 62,389 mutants and covering 83% of the nuclear protein-coding genes. A genome-wide survey of genes required for photosynthesis identified 303 candidate genes. Of these, 21 of the 43 high-confidence genes were newly identified and relevant for photosynthesis.

5. Application of the transposon insertion site sequencing method in microalgae

Microalgae are considered significant renewable biological resources as the mainstay of photosynthesis on Earth. Certain algae have high biomass, short growth cycles, are easy to culture, and have a high content of valuable substances. Using transposons to insert mutations into microalgae genes and sequencing insertion sites to understand insertion locations and genes to determine microalgae gene functions and between-gene interrelationships are standard methods in this biological group.

High-throughput sequencing will be the primary method for studying gene function in the future, based on the current research trend of microalgae. A large amount of transposon insertion site data will be obtained by high-throughput sequencing concerning gene function annotation or gene fitness to obtain gene expression in different growth environments. This becomes more of a need for methods that allow high-throughput determination of transposon insertion sites. However, using high-throughput sequencing methods becomes less necessary to determine transposon insertion sites for individual mutants with obvious phenotypes, especially from the point of view of costs. It is simple and easy to control costs using enzyme digestion and multiple primer amplification methods (Figure 9).

Fauser et al. (2022) determined the insertion site of each transposon by high-throughput sequencing using random transposon insertion into the genome of the model organism *C. reinhardtii*. These authors determined the phenotype of over 58,000 mutants by screening them under more than 121 different environmental growth conditions and chemical treatments. Fifty-nine percent of the genes in *C. reinhardtii* were represented by transposon insertion mutants that exhibited at least one phenotype. This is the most complete and comprehensive library of eukaryotic microalgal mutants known, providing a basis for the function of thousands of genes in *C. reinhardtii*. Previously, functionally unknown genes could be identified based on their functions, including DNA repair, photosynthesis, CO_2 concentration mechanisms, and ciliogenesis.

Broddrick et al. (2016) used random transposon insertion to create a mutant library of prokaryotic microalgae in PCC 7942. These researchers identified genes essential for PCC 7942 under specific growth conditions through changes in the fitness of individual genes under different growth conditions. The genome-scale metabolic model of PCC 7942 was revised to produce a highly accurate metabolic model. Some previously unknown metabolic features of PCC 7942 were identified, including the nonessential nature of the TCA cycle.

6. Conclusion and future perspective

Transposon insertion marker DNA has become an essential tool for studying the functional genomics of organisms. A large number of DNA insertion lines and important mutations have been created in microalgae using this approach, which is necessary to determine the genomic sequence on either side of the insertion marker to identify genes tagged by transposon insertion. However, the sequences of the tagged genes cannot be obtained simply by conventional PCR reactions, which require a specific experimental design and technical methodological modifications. Current sequencing methods have distinctive characteristics and different problems. The reverse PCR and plasmid rescue methods are simple and operationally uncomplicated, with easily controllable costs but lower success rates. The special enzyme digestion method is more specific but due to the restrictive enzyme digestion sites, resulting in insufficient applicability. TAIL PCR and RACE technological steps have higher success rates than previous methods (Chen N. et al., 2016; Tan et al., 2019). Still, pre-processing LAM-PCR and Tn-Seq have high success rates and are especially suitable for high-throughput sequencing and establishing mutant libraries. However, the cost of sequencing a single mutant is high, and the amount of invalid data during sequencing is vast, which may be due to the lack of sufficient specificity of the sequencing primer used or the insufficient screening capacity of the available equipment for large amounts of data. These invalid data can be filtered in subsequent data processing, and generally will not cause errors in subsequent analysis and target selection. Therefore, we need to consider several factors when arranging sequencing experiments (e.g., experimental conditions, experimental schedule, cost, and the combination of multiple methods for sequencing and validation) to make reasonable and flexible experimental arrangements (Table 2).

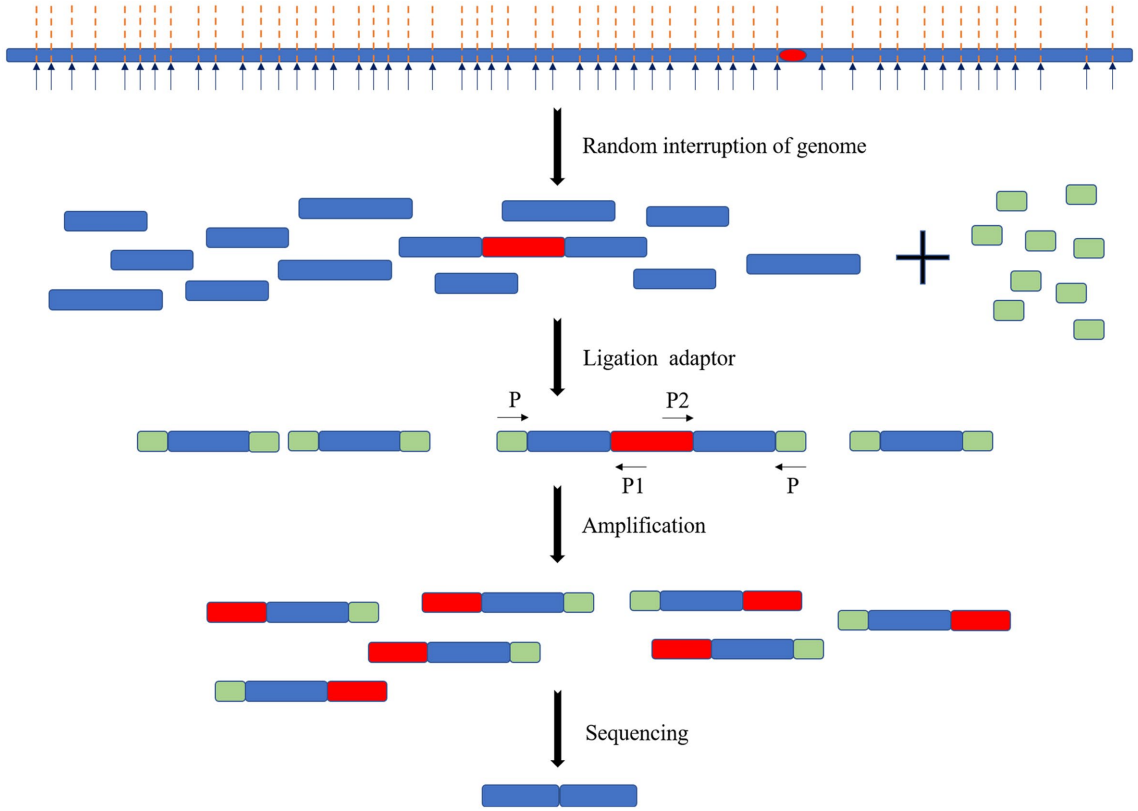

FIGURE 8
Schematic diagram of the Tn-Seq principle. The genome of the transposon insert was randomly interrupted. The splice sequence is added, and primers are designed with the specific sequences of the splice sequence and transposon for amplification and sequencing.

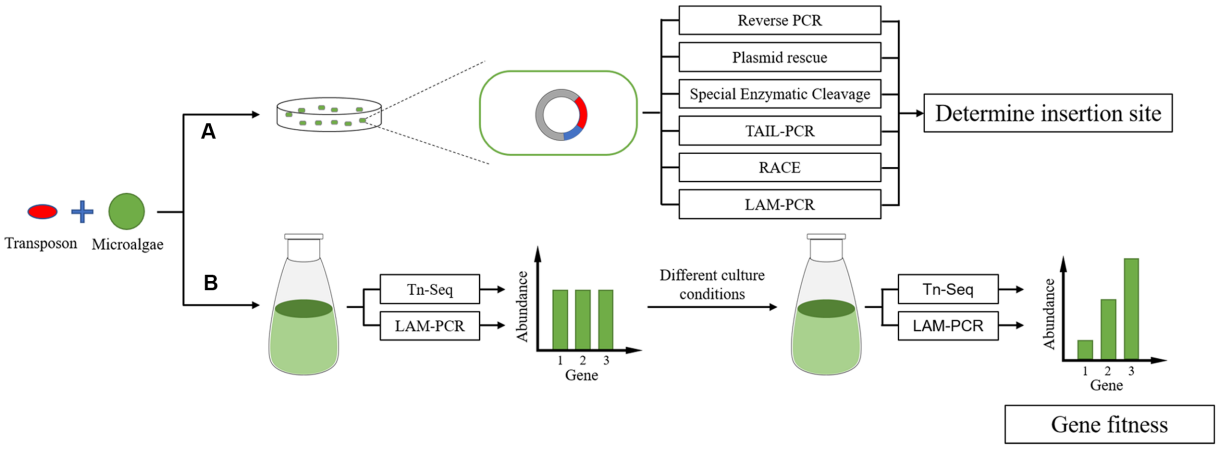

FIGURE 9
Transposon insertion site sequencing in microalgae. **(A)** Individual mutants with obvious phenotypes and research values are often sequenced using simple methods for equipment and operation. **(B)** High-throughput microalgae sequencing to establish a transposon insertion mutant library.

The main future development direction will be improving the success rate and cost control to solve the problem of flanking sequencing after transposon insertion. It is currently difficult for all sequencing methods to reach a 90% success rate. A large amount of invalid data needs to be processed, even for high-throughput sequencing, which invariably raises the technical threshold and labor costs of the equipment. It is necessary to improve the transposon and sequencing methods to solve these problems. First, the transposon can be modified while retaining random insertion ability. The transposon itself should be able to carry a more easily identifiable tag, reducing the misoperation of devices in the amplifying process of target sequences and sequencing reads. Second, with technological and equipment updates, the accurate sequencing method is constantly updated, and sequencing costs decrease.

TABLE 2 Comparison of sequencing methods.

Method		Advantages	Disadvantages
Enzymatic digestion	Reverse PCR	Simple operation, low cost.	The sequencing results of mutation sites are unstable and not suitable for high-throughput sequencing.
	Plasmid rescue	The results are not prone to false positives.	Not suitable for high-throughput sequencing.
	Specific enzymatic cleavage	The method is simple and supports high-throughput sequencing.	The short sequence used for sequencing leads to unstable results.
Multiple primer amplification method	TAIL PCR	Strong specificity, high efficiency.	Not suitable for high-throughput sequencing.
	RACE	Strong specificity, high efficiency.	Not suitable for high-throughput sequencing.
	LAM-PCR	Strong specificity, high success rate, suitable for high-throughput sequencing.	High cost.
Tn-Seq		Strong specificity, high success rate, suitable for high-throughput sequencing.	Long test cycle, heavy data processing workload.

In conclusion, determining bipartite sequences after transposon insertion will be increasingly accessible, fast, and inexpensive with the development of various sequencing methods and transposon technologies. Applying the latest gene function research methods to microalgae can facilitate effective transformation and make them more excellent engineering microalgal chassis cells. Consequently, they can contribute better to human food, energy, and environmental protection.

Author contributions

XH and QW conceptualized the idea for manuscript. XH, YF, CM, and HC drafted the manuscript. QW evaluated the manuscript and improved the content. All authors contributed to the article and approved the submitted version.

Funding

This work was supported jointly by the National Key R&D Program of China (2021YFA0909600), the National Natural Science Foundation of China (32170138 and 31870041), the Natural Science Foundation of Henan Province (212300410024), the Program for Innovative Research Team (in Science and Technology) in University of Henan Province (22IRTSTHN024), and the 111 Project (#D16014).

Acknowledgments

The authors would like to express gratitude to EditSprings (https://www.editsprings.cn) for the expert linguistic services provided.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

References

Beaumont, M., Tran, R., Vera, G., Niedrist, D., Rousset, A., Pierre, R., et al. (2021). Hydrogel-forming algae polysaccharides: from seaweed to biomedical applications. *Biomacromolecules* 22, 1027–1052. doi: 10.1021/acs.biomac.0c01406

Bigelow, T. A., Xu, J., Stessman, D. J., Yao, L., Spalding, M. H., and Wang, T. (2014). Lysis of *Chlamydomonas reinhardtii* by high-intensity focused ultrasound as a function of exposure time. *Ultrason. Sonochem.* 21, 1258–1264. doi: 10.1016/j.ultsonch.2013.11.014

Broddrick, J. T., Rubin, B. E., Welkie, D. G., Du, N., Mih, N., Diamond, S., et al. (2016). Unique attributes of cyanobacterial metabolism revealed by improved genome-scale metabolic modeling and essential gene analysis. *Proc. Natl. Acad. Sci.* 113, E8344–E8353. doi: 10.1073/pnas.1613446113

Cabanelas, I. T., Ruiz, J., Arbib, Z., Chinalia, F. A., Garrido-Perez, C., Rogalla, F., et al. (2013). Comparing the use of different domestic wastewaters for coupling microalgal production and nutrient removal. *Bioresour. Technol.* 131, 429–436. doi: 10.1016/j.biortech.2012.12.152

Carette, J. E., Guimaraes, C. P., Wuethrich, I., Blomen, V. A., Varadarajan, M., Sun, C., et al. (2011). Global gene disruption in human cells to assign genes to phenotypes by deep sequencing. *Nat. Biotechnol.* 29, 542–546. doi: 10.1038/nbt.1857

Chen, N., Wang, W. M., and Wang, H. L. (2016). An efficient full-length cDNA amplification strategy based on bioinformatics technology and multiplexed PCR methods. *Sci. Rep.* 6:19420. doi: 10.1038/srep19420

Chen, W., Zhang, S., Rong, J., Li, X., Chen, H., He, C., et al. (2016). Effective biological DeNOx of industrial flue gas by the mixotrophic cultivation of an oil-producing green alga chlorella sp. C2. *Environ. Sci. Technol.* 50, 1620–1627. doi: 10.1021/acs.est.5b04696

Choi, K. H., and Kim, K. J. (2009). Applications of transposon-based gene delivery system in bacteria. *Korean Soc. Microbiol. Biotechnol.* 19, 217–228. doi: 10.4014/jmb.0811.669

Chutia, S. J., Bora, G., Kumar, M., Nath, R. J., BS, Y., Dihingia, P., et al. (2020). Recent developments in RACE-PCR for the full-length cDNA identification. *J. Entomol. Zool. Stud.* 8, 444–449.

Damasceno, J. D., Beverley, S. M., and Tosi, L. R. O. (2010). A transposon toolkit for gene transfer and mutagenesis in protozoan parasites. *Genetica* 138, 301–311. doi: 10.1007/s10709-009-9406-7

Fauser, F., Vilarrasa-Blasi, J., Onishi, M., Ramundo, S., Patena, W., Millican, M., et al. (2022). Systematic characterization of gene function in the photosynthetic alga

Chlamydomonas reinhardtii. *Nat. Genet.* 54, 705–714. doi: 10.1038/s41588-022-01052-9

Feng, D.-S., Wang, H.-G., Zhang, X.-S., Kong, L.-R., Tian, J.-C., and Li, X.-F. (2008). Using an inverse PCR method to clone the wheat cytokinin oxidase/dehydrogenase gene TaCKX1. *Plant Mol. Biol. Report.* 26, 143–155. doi: 10.1007/s11105-008-0033-8

Frigon, J.-C., Matteau-Lebrun, F., Hamani Abdou, R., McGinn, P. J., O'Leary, S. J. B., and Guiot, S. R. (2013). Screening microalgae strains for their productivity in methane following anaerobic digestion. *Appl. Energy* 108, 100–107. doi: 10.1016/j.apenergy.2013.02.051

Fu, Y., Chen, T., Chen, S. H. Y., Liu, B., Sun, P., Sun, H., et al. (2021). The potentials and challenges of using microalgae as an ingredient to produce meat analogues. *Trends Food Sci. Technol.* 112, 188–200. doi: 10.1016/j.tifs.2021.03.050

Gabriel, R., Kutschera, I., Bartholomae, C. C., von Kalle, C., and Schmidt, M. (2014). Linear amplification mediated PCR – localization of genetic elements and characterization of unknown flanking DNA. *JoVE* 88:e51543. doi: 10.3791/51543

Gierl, A., and Saedler, H. (1992). Plant-transposable elements and gene tagging. *Plant Mol. Biol.* 19, 39–49. doi: 10.1007/bf00015605

Gomez, C., Escudero, R., Morales, M. M., Figueroa, F. L., Fernandez-Sevilla, J. M., and Acien, F. G. (2013). Use of secondary-treated wastewater for the production of Muriellopsis sp. *Appl. Microbiol. Biotechnol.* 97, 2239–2249. doi: 10.1007/s00253-012-4634-7

Gutiérrez, L. G., Abelleyro, M. M., Ruiz, M. S., Anchordoqui, M. S., Freitas, J., Bianchini, M., et al. (2021). Development of an inverse-PCR approach for characterization of the major BCR-ABL1 breakpoint sequences on genomic DNA: proof of concept. *Clin. Chem. Lab. Med.* 59, e449–e453. doi: 10.1515/cclm-2020-1482

Hu, J., Li, T., Xu, W., Zhan, J., Chen, H., He, C., et al. (2017). Small antisense RNA RblR positively regulates RuBisCo in *Synechocystis* sp. PCC 6803. *Front. Microbiol.* 8:231. doi: 10.3389/fmicb.2017.00231

Huang, A. M., Rehm, E. J., and Rubin, G. M. (2009). Recovery of DNA sequences flanking P-element insertions in *Drosophila*: inverse PCR and plasmid rescue. *Cold Spring Harb. Protoc.* 2009:pdb.prot5199. doi: 10.1101/pdb.prot5199

Kemppainen, M., Duplessis, S., Martin, F., and Pardo, A. G. (2008). T-DNA insertion, plasmid rescue and integration analysis in the model mycorrhizal fungus *Laccaria bicolor*. *Microb. Biotechnol.* 1, 258–269. doi: 10.1111/j.1751-7915.2008.00029.x

Li, N., Jin, K., Bai, Y., Fu, H., Liu, L., and Liu, B. (2020). Tn5 transposase applied in genomics research. *Int. J. Mol. Sci.* 21:8329. doi: 10.3390/ijms21218329

Li, X., Patena, W., Fauser, F., Jinkerson, R. E., Saroussi, S., Meyer, M. T., et al. (2019). A genome-wide algal mutant library and functional screen identifies genes required for eukaryotic photosynthesis. *Nat. Genet.* 51, 627–635. doi: 10.1038/s41588-019-0370-6

Li, X., Xue, C., Chen, H., Zhang, H., and Wang, Q. (2022). Small antisense RNA ThfR positively regulates Thf1 in *Synechocystis* sp. PCC 6803. *J. Plant Physiol.* 271:153642. doi: 10.1016/j.jplph.2022.153642

Liu, Y. G., and Chen, Y. (2007). High-efficiency thermal asymmetric interlaced PCR for amplification of unknown flanking sequences. *Biotechniques* 43, 649–650. doi: 10.2144/000112601

Liu, Y.-G., and Huang, N. (1998). Efficient amplification of insert end sequences from bacterial artificial chromosome clones by thermal asymmetric interlaced PCR. *Plant Mol. Biol. Report.* 16, 175–181. doi: 10.1023/A:1007420918645

Liu, R., Li, S., Tu, Y., Hao, X., and Qiu, F. (2022). Recovery of value-added products by mining microalgae. *J. Environ. Manag.* 307:114512. doi: 10.1016/j.jenvman.2022.114512

Liu, Y.-G., Mitsukawa, N., Oosumi, T., and Whittier, R. F. (1995). Efficient isolation and mapping of *Arabidopsis thaliana* T-DNA insert junctions by thermal asymmetric interlaced PCR. *Plant J.* 8, 457–463. doi: 10.1046/j.1365-313X.1995.08030457.x

Liu, Y.-G., and Wittier, R. F. (1995). Thermal asymmetric interlaced PCR: automatable amplification and sequencing of insert end fragments from PI and YAC clones for chromosome walking. *Genomics* 25, 674–681. doi: 10.1016/0888-7543(95)80010-J

Matsumura, H., Reich, S., Ito, A., Saitoh, H., Kamoun, S., Winter, P., et al. (2003). Gene expression analysis of plant host–pathogen interactions by SuperSAGE. *Proc. Natl. Acad. Sci.* 100, 15718–15723. doi: 10.1073/pnas.2536670100

Meslet-Cladière, L., and Vallon, O. (2012). A new method to identify flanking sequence tags in chlamydomonas using 3'-RACE. *Plant Methods* 8:21. doi: 10.1186/1746-4811-8-21

Ng, I. S., Keskin, B. B., and Tan, S. I. (2020). A critical review of genome editing and synthetic biology applications in metabolic engineering of microalgae and cyanobacteria. *Biotechnol. J.* 15:e1900228. doi: 10.1002/biot.201900228

Ng, P., Wei, C. L., Sung, W. K., Chiu, K. P., Lipovich, L., Ang, C. C., et al. (2005). Gene identification signature (GIS) analysis for transcriptome characterization and genome annotation. *Nat. Methods* 2, 105–111. doi: 10.1038/nmeth733

Oranab, S., Ghaffar, A., Kiran, S., Yameen, M., Munir, B., Zulfiqar, S., et al. (2021). Molecular characterization and expression of cyclic nucleotide gated ion channels 19 and 20 in Arabidopsis thaliana for their potential role in salt stress. *Saudi J. Biol. Sci.* 28, 5800–5807. doi: 10.1016/j.sjbs.2021.06.027

Packer, M. A., Harris, G. C., and Adams, S. L. (2016). "Food and feed applications of algae" in *Algae Biotechnology* eds. F. Bux and Y. Chisti (Cham: Springer), 217–247.

Passmore, L. A., and Coller, J. (2021). Roles of mRNA poly(A) tails in regulation of eukaryotic gene expression. *Nat. Rev. Mol. Cell Biol.* 23, 93–106. doi: 10.1038/s41580-021-00417-y

Rubin, B. E., Wetmore, K. M., Price, M. N., Diamond, S., Shultzaberger, R. K., Lowe, L. C., et al. (2015). The essential gene set of a photosynthetic organism. *Proc. Natl. Acad. Sci.* 112, E6634–E6643. doi: 10.1073/pnas.1519220112

Schmidt, M., Schwarzwaelder, K., Bartholomae, C., Zaoui, K., Ball, C., Pilz, I., et al. (2007). High-resolution insertion-site analysis by linear amplification-mediated PCR (LAM-PCR). *Nat. Methods* 4, 1051–1057. doi: 10.1038/nmeth1103

Scieszka, S., and Klewicka, E. (2019). Algae in food: a general review. *Crit. Rev. Food Sci. Nutr.* 59, 3538–3547. doi: 10.1080/10408398.2018.1496319

Shapiro, J. A. (2010). Mobile DNA and evolution in the 21st century. *Mob. DNA* 1:4. doi: 10.1186/1759-8753-1-4

Singh, U. B., and Ahluwalia, A. S. (2012). Microalgae: a promising tool for carbon sequestration. *Mitig. Adapt. Strateg. Glob. Chang.* 18, 73–95. doi: 10.1007/s11027-012-9393-3

Su, Z., Saha, S., Paulsen, T., Kumar, P., and Dutta, A. (2021). ATAC-Seq-based identification of extrachromosomal circular DNA in mammalian cells and its validation using inverse PCR and FISH. *Bio Protoc.* 11:e4003. doi: 10.21769/BioProtoc.4003

Sun, Z., Chen, H., and Wang, Q. (2022). From CO_2 to value-added products—carbon neutral microalgal green biomanufacturing. *Synth. Biol. J.* 3, 953–965. doi: 10.12211/2096-8280.2022-023

Tan, J., Gong, Q., Yu, S., Hou, Y., Zeng, D., Zhu, Q., et al. (2019). A modified high-efficiency thermal asymmetric interlaced PCR method for amplifying long unknown flanking sequences. *J. Genet. Genomics* 46, 363–366. doi: 10.1016/j.jgg.2019.05.002

Tsurumaru, H., Yamakawa, T., Tanaka, M., and Sakai, M. (2008). The efficient strategy of plasmid rescue from Tn5 mutants derived from *Bradyrhizobium japonicum* Is-1, based on whole genome sequence information of strain USDA110. *J. Fac. Agric. Kyushu Univ.* 53, 27–31. doi: 10.5109/10065

van Opijnen, T., and Camilli, A. (2013). Transposon insertion sequencing: a new tool for systems-level analysis of microorganisms. *Nat. Rev. Microbiol.* 11, 435–442. doi: 10.1038/nrmicro3033

Vidyashankar, S., VenuGopal, K. S., Chauhan, V. S., Muthukumar, S. P., and Sarada, R. (2014). Characterisation of defatted *Scenedesmus dimorphus* algal biomass as animal feed. *J. Appl. Phycol.* 27, 1871–1879. doi: 10.1007/s10811-014-0498-9

Wang, J., Bi, X., Chen, W., Zhao, Q., Yang, J., Tong, X., et al. (2021). Identification of the insertion site of transgenic DNA based on cyclization of the target gene with the flanking sequence and nested inverse PCR. *Talanta Open* 3:100033. doi: 10.1016/j.talo.2021.100033

Wang, G. H., Xiao, J. H., Xiong, T. L., Li, Z., Murphy, R. W., and Huang, D. W. (2013). High-efficiency thermal asymmetric interlaced PCR (hiTAIL-PCR) for determination of a highly degenerated prophage WO genome in a Wolbachia strain infecting a fig wasp species. *Appl. Environ. Microbiol.* 79, 7476–7481. doi: 10.1128/AEM.02261-13

Wetmore, K. M., Price, M. N., Waters, R. J., Lamson, J. S., He, J., Hoover, C. A., et al. (2015). Rapid quantification of mutant fitness in diverse bacteria by sequencing randomly bar-coded transposons. *mBio* 6, e00306–e00315. doi: 10.1128/mBio.00306-15

Yuan, L., Li, Y., Zhang, X., Guo, W., Che, Y., and Chen, G. (2009). Quick identification of pathogenicity-related genes in *Xanthomonas oryzae* pv. oryzicola by thermal asymmetric interlaced PCR(TAIL-PCR) and Tn5 transposon rescue. *J. Agricult. Biotechnol.* 17, 1089–1095. doi: 10.3969/j.issn.1674-7968.2009.06.023

Zhang, Y., Fan, X., Yang, Z., Wang, H., Yang, D., and Guo, R. (2012). Characterization of H2 photoproduction by a new marine green alga, *Platymonas helgolandica* var. *tsingtaoensis*. *Appl. Energy* 92, 38–43. doi: 10.1016/j.apenergy.2011.09.044

Zhang, R., Patena, W., Armbruster, U., Gang, S. S., Blum, S. R., and Jonikas, M. C. (2014). High-throughput genotyping of green algal mutants reveals random distribution of mutagenic insertion sites and endonucleolytic cleavage of transforming DNA. *Plant Cell* 26, 1398–1409. doi: 10.1105/tpc.114.124099

frontiers | Frontiers in Microbiology

TYPE Review
PUBLISHED 14 February 2023
DOI 10.3389/fmicb.2023.1126032

OPEN ACCESS

EDITED BY
Martin Hagemann,
University of Rostock,
Germany

REVIEWED BY
Guodong Luan,
Qingdao Institute of Bioenergy and Bioprocess Technology (CAS), China
Shailendra Pratap Singh,
Institute of Science,
Banaras Hindu University,
India

*CORRESPONDENCE
María Santos-Merino
✉ santosm7@msu.edu
Daniel C. Ducat
✉ ducatdan@msu.edu

†These authors have contributed equally to this work

SPECIALTY SECTION
This article was submitted to Microbiotechnology, a section of the journal Frontiers in Microbiology

RECEIVED 17 December 2022
ACCEPTED 24 January 2023
PUBLISHED 14 February 2023

CITATION
Santos-Merino M, Yun L and Ducat DC (2023) Cyanobacteria as cell factories for the photosynthetic production of sucrose. *Front. Microbiol.* 14:1126032. doi: 10.3389/fmicb.2023.1126032

Cyanobacteria as cell factories for the photosynthetic production of sucrose

María Santos-Merino[1*†], Lisa Yun[1,2†] and Daniel C. Ducat[1,2*]

[1]MSU-DOE Plant Research Laboratory, Michigan State University, East Lansing, MI, United States, [2]Department of Biochemistry and Molecular Biology, Michigan State University, East Lansing, MI, United States

Biofuels and other biologically manufactured sustainable goods are growing in popularity and demand. Carbohydrate feedstocks required for industrial fermentation processes have traditionally been supplied by plant biomass, but the large quantities required to produce replacement commodity products may prevent the long-term feasibility of this approach without alternative strategies to produce sugar feedstocks. Cyanobacteria are under consideration as potential candidates for sustainable production of carbohydrate feedstocks, with potentially lower land and water requirements relative to plants. Several cyanobacterial strains have been genetically engineered to export significant quantities of sugars, especially sucrose. Sucrose is not only naturally synthesized and accumulated by cyanobacteria as a compatible solute to tolerate high salt environments, but also an easily fermentable disaccharide used by many heterotrophic bacteria as a carbon source. In this review, we provide a comprehensive summary of the current knowledge of the endogenous cyanobacterial sucrose synthesis and degradation pathways. We also summarize genetic modifications that have been found to increase sucrose production and secretion. Finally, we consider the current state of synthetic microbial consortia that rely on sugar-secreting cyanobacterial strains, which are co-cultivated alongside heterotrophic microbes able to directly convert the sugars into higher-value compounds (e.g., polyhydroxybutyrates, 3-hydroxypropionic acid, or dyes) in a single-pot reaction. We summarize recent advances reported in such cyanobacteria/heterotroph co-cultivation strategies and provide a perspective on future developments that are likely required to realize their bioindustrial potential.

KEYWORDS

cyanobacteria, sucrose metabolism, carbohydrate feedstocks, osmoprotection, co-cultures

1. Introduction

Product generation through heterotrophic microbial fermentation has been successfully used as an alternative approach to classical chemical processes using petroleum-based feedstocks (Blombach et al., 2022). However, bioindustrial chemical production by bacterial fermentation is still not economically competitive for many commodity products due in part to the high costs associated to the carbon substrates used for these organisms (Lee et al., 2022). Extensive research efforts have been expended to identify new plant species or to improve biomass processing technologies and increase the yield of fermentable sugars from plant feedstocks (e.g., improving carbohydrate recovery from cellulosic materials; Sun et al., 2022) and to overcome other land-use problems of plant-based feedstocks (Das and Gundimeda, 2022). Thus, there is an increased interest on the search for alternative, economical and environmentally sustainable sources as carbohydrate feedstocks.

Cyanobacteria and microalgae have attracted more attention in the last few years as an alternative supply for carbohydrates to support industrial fermentative processes (Hays and Ducat, 2015;

Santos-Merino et al., 2019). In comparison with plants, cyanobacteria and algae can tolerate many water supplies that are unsuitable for agriculture (Santos-Merino et al., 2019; Catone et al., 2021), reducing their competition with food crops for the limited supply of arable land and freshwater. Microalgae and cyanobacteria are generally easier to manipulate genetically, have rapid division times, and can achieve higher efficiencies of solar energy capture and conversion (Santos-Merino et al., 2019). Relative to microalgae that tend to store excess carbon in the form of lipids or starch (Scott et al., 2010), cyanobacteria normally accumulate carbon reserves in polysaccharides and frequently sucrose as a compatible solute (osmolyte) in high-salt environments or under other abiotic stress (Klähn and Hagemann, 2011; Kirsch et al., 2019; Sanz Smachetti et al., 2020). Sucrose metabolism and its regulation has been amply studied in cyanobacteria (Kolman et al., 2015; Kirsch et al., 2019), while the activity and regulation of sucrose metabolism factors has received less attention in microalgae (Radakovits et al., 2010; Hagemann, 2016). The increasing knowledge on the synthesis and regulation of sucrose not only improves our understanding of these pathways but will also be useful for genetically engineering them for future biotechnological applications.

A number of cyanobacterial species have been effectively engineered to produce and secrete large amounts of sucrose by taking advantage of cyanobacterial sucrose biosynthesis pathways and heterologous co-expression of sucrose permease (CscB, Ducat et al., 2012; Du et al., 2013; Abramson et al., 2016; Kirsch et al., 2018; Lin P. C. et al., 2020) to export sucrose from the cell. In addition to batch cultures, there are increasing examples of real-time conversion of the carbohydrate feedstock through the direct co-culture of microbial partner strains that metabolize the secreted bacterial sucrose to higher-value products (Smith and Francis, 2016; Hays et al., 2017; Löwe et al., 2017; Weiss et al., 2017; Fedeson et al., 2020; Hobmeier et al., 2020; Zhang et al., 2020; Ma et al., 2022; Kratzl et al., 2023), potentially bypassing the costly processes of purifying and concentrating sucrose (Radakovits et al., 2010). However, to further use these synthetic light-driven microbial consortia in industrial applications, a number of challenges need to be overcome, such as long-term production stability, vulnerability to invasion by opportunistic microbial or viral contaminants, and imbalances in attributes of consortia that can contribute to inefficiencies (Hays and Ducat, 2015; Gao et al., 2022).

This review focuses on the current knowledge of the sucrose synthesis and degradation pathways in cyanobacteria as well as the list of genetic modifications in sucrose metabolic pathways that have been found to increase the production and secretion of this sugar. While other sugars can be produced phototrophically from cyanobacteria (e.g., glucose, fructose, or polysaccharides; Niederholtmeyer et al., 2010; Arias et al., 2021), sucrose has been the highest yielding carbohydrate reported and is the main focus of this review. We highlight some unresolved questions for additional study on fundamental cyanobacterial sucrose metabolism and the utilization of these pathways for bioproduction. Finally, we examine the current state of synthetic microbial consortia that capitalize upon the carbon fixation that photoautotrophs like cyanobacteria are uniquely able to provide.

2. Cyanobacterial sucrose metabolism

2.1. Sucrose biosynthesis pathway

Sucrose is a disaccharide [α-d-glucopyranosyl (1→2) β-d-fructofuranoside], whose synthesis pathway appears to be nearly universal among cyanobacteria, as predicted by the presence of sucrose synthesis genes in most of the known genome sequences available so far (Kolman et al., 2015; Kirsch et al., 2019; Supplementary Table S1). Sucrose synthesis lies close to the core of central carbon metabolism, with substrates directly derived from the Calvin-Benson-Bassham (CBB) cycle and its immediate downstream products (Figure 1). The light reactions of photosynthesis generate NADPH and ATP, which are used in the CBB to fix CO_2 and yield glyceraldehyde-3-phosphate (GAP). GAP can be interchangeably converted to dihydroxyacetone phosphate (DHAP), and the condensation of GAP and DHAP through the activity of the enzyme fructose 1,6-bisphosphate aldolase (FBA), leads to the formation of fructose 1,6-bisphosphate (FBP). FBP is then further transformed into other hexose phosphates, such as fructose 6-phophate (F6P) and glucose 6-phosphate (G6P). G6P can be used to form nucleotide sugars such as uridine diphosphate glucose (UDP-Gluc).

Sucrose is most commonly synthesized from these CBB products in a two-step reaction by the sequential activity of two enzymes, sucrose phosphate synthase (SPS) and sucrose phosphate phosphatase (SPP; Figure 1). NDP-Gluc is combined with F6P to form sucrose 6-phosphate (S6P) in a reaction catalyzed by SPS. S6P is then dephosphorylated by SPP to sucrose, concluding the sucrose biosynthesis pathway. The rapid irreversible hydrolysis of S6P by a specific and high-activity SPP drives the reversible reaction catalyzed by SPS towards the direction of sucrose synthesis, even at low substrate concentrations (Lunn and ap Rees, 1990). An alternative route for sucrose synthesis is catalyzed by the enzyme sucrose synthase (SuS), which binds UDP/ADP-Gluc with fructose to produce sucrose (Porchia et al., 1999; Figure 1). While SuS is able to catalyze the synthesis of sucrose, cellular energetics are such that SuS is thought to be solely involved in sucrose cleavage *in vivo* (Curatti et al., 2002). SuS is ubiquitous across plant species, in contrast with cyanobacteria, where its occurrence is not widespread (Salerno and Curatti, 2003).

2.1.1. Sucrose phosphate synthase

SPS catalyzes the first step in the pathway of sucrose synthesis by transferring a glycosyl group from an activated donor sugar, such as UDP-Gluc, to a sugar acceptor F6P, resulting in the formation of UDP and S6P (Figure 1). SPS (EC 2.4.1.14) is a UDP–glucose: d-fructose-6-phosphate 2-α-d-glucosyltransferase belonging to the GT-B (glucosyltransferase fold B) type glucosyltransferase family and its secondary structure consists of two distinct Rossmann-fold domains (super-secondary structures composed of consecutive alternating β-strands and α-helices that form a layer of β-sheet with one/two layer/s of α-helices) - a sugar acceptor domain (N-terminal "A-domain") and a sugar donor domain (C-terminal "B-domain"; Chua et al., 2008; Lairson et al., 2008). In a recent report, the structure of the SPS from *Thermosynechococcus elongatus* was resolved, showing that this enzyme has 16 α-helices and 14 β-sheets, with UDP and S6P bound at the interface of the aforementioned A-and B-domains (Li et al., 2020). Whereas plant SPSs are specific for UDP-Gluc, cyanobacterial SPSs are not, and can accept other NDP-Gluc forms as substrates, such as ADP-Gluc and GDP-Gluc (Curatti et al., 1998; Lunn et al., 1999; Gibson et al., 2002). Another difference between cyanobacterial and plant SPSs is that the activity of the latter is regulated by light–dark modulation *via* reversible phosphorylation (Winter and Huber, 2000; Li et al., 2020).

In cyanobacteria, the glucosyltransferase domain (GTD) of SPS contains two motifs that are highly conserved across glucosyl-transferase family enzymes (Figure 2A). Motif I (G-X_5-GGQ-X_2-Y-X_2-EL) is located in the N-terminus of SPS and has been hypothesized to include residues

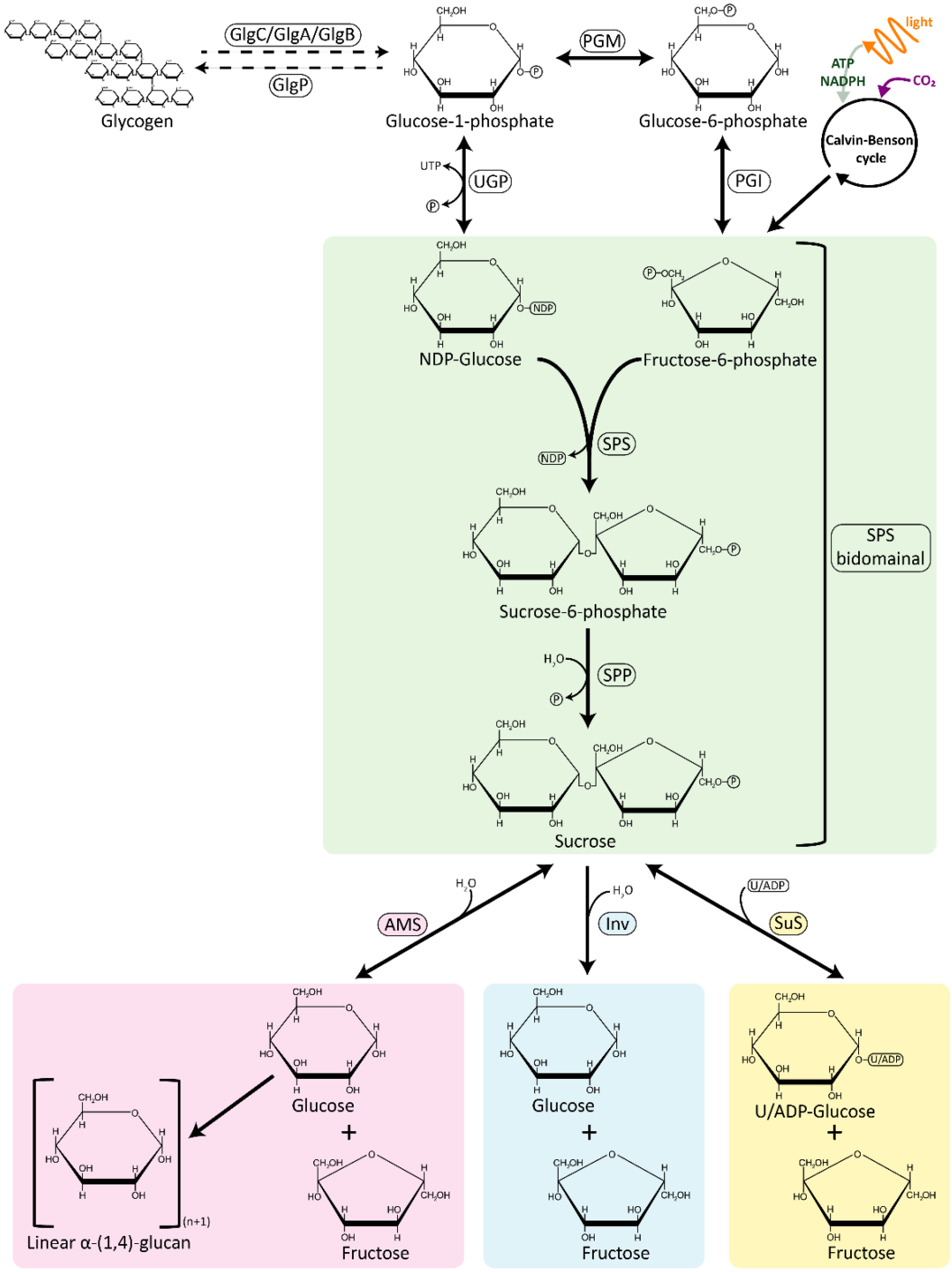

FIGURE 1
An overview of sucrose synthesis and degradation pathways in cyanobacteria. The sucrose synthesis pathway is represented in green; the degradation pathways are represented in pink, blue, and yellow. AMS, amylosucrase; GlgA, glycogen synthase; GlgB, glycogen branching enzyme; GlgC, ADP-glucose pyrophosphorylase; GlgP, glycogen phosphorylase; INV, invertase; PGI, phosphoglucose isomerase; PGM, phosphoglucomutase; SPP, sucrose phosphate phosphatase; SPS, sucrose phosphate synthase; SuS, sucrose synthase; UGP, UDP-glucose pyrophosphorylase.

necessary for defining the F6P binding site (Figure 2C, left panel; Ma et al., 2020). Motif II (E-X_7-E) is highly conserved within the C-terminus of SPS and SuS enzymes; its flanking Glu residues play a catalytic role in the reaction by binding to UDP-Gluc (Figure 2C, left panel; Cumino et al., 2002; Ma et al., 2020; Kurniah et al., 2021). The first Glu residue of E-X_7-E may function as the nucleophile, whereas the second Glu may function as the general acid/base catalyst (Cid et al., 2000). Both Glu residues in motif II are important for SPS activity, as demonstrated by point mutants in the GTD domain (E356A and E364A in SPS_{7942}) of the bidomainal SPS encoded in *Synechococcus elongatus* PCC 7942 (SPS_{7942}; see discussion below on bidomainal proteins) that disrupted sucrose synthesis, specifically preventing S6P formation (Liang et al., 2020).

2.1.2. Sucrose phosphate phosphatase

The reversible reaction catalyzed by SPS is followed by the irreversible dephosphorylation of S6P to sucrose by SPP (Figure 1). SPP

10.3389/fmicb.2023.1126032

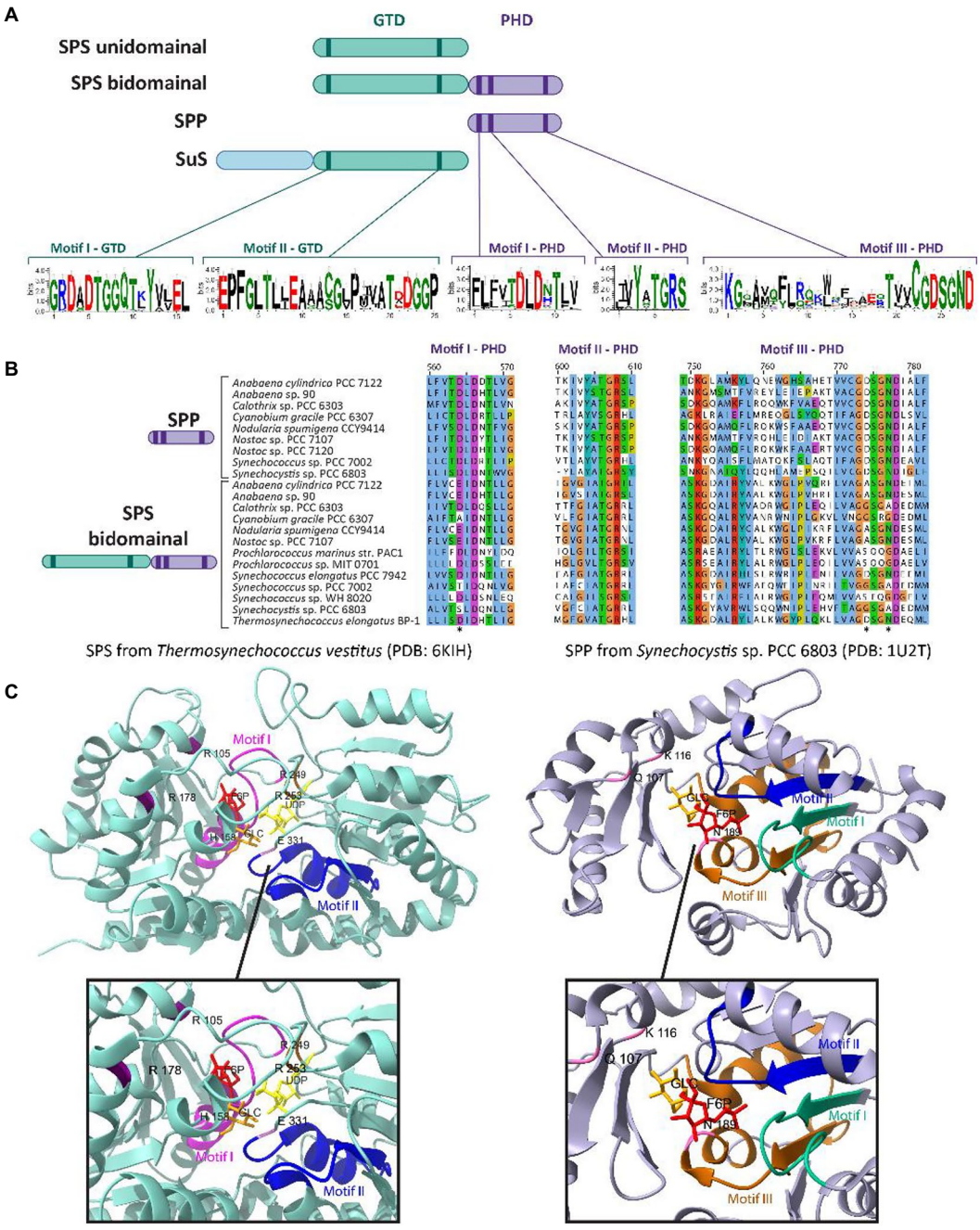

FIGURE 2
Conserved domains and motifs of SPS, SPP, and SuS among cyanobacterial species. **(A)** Schematic cartoon representing the domainal arrangements and the motifs present in SPS, SPP, and SuS. The glucosyl-transferase domain (GTD) is represented in green, while the phosphohydrolase domain (PHD) is represented in purple. The extended N-terminal domain found in SuS is represented in blue. GTD and PHD domains contain two and three conserved motifs, respectively. Logos for these conserved motifs were obtained using the WebLogo server (Crooks et al., 2004). **(B)** Multiple sequence alignment analysis of the deduced amino acid sequences for the three motifs present in the PHD domain. The alignment was performed using MEGA X (Kumar et al., 2018) and visualized with the Jalview multiple sequence alignment editor using the color scheme from ClustalX (Waterhouse et al., 2009). The asterisks indicate conserved residues that are mutated in the PHD domain in the sequences of cyanobacterial SPS bidomainal proteins relative to unidomainal homologs. **(C)** Crystal structure of the SPS from *T. vestivus* (PDB: 6KIH) and the SPP from *Synechocystis* sp. PCC 6803 (PDB: 1U2T) highlighting the motifs indicated in **(A)** and the residues involved in binding to their respective substrates (top) and a zoom-in of the catalytic centers of each enzyme (bottom). In the SPS (left panel), the residues R105, R178, R249 and R253 stabilize phosphate group of S6P; R249 and R253 stabilize the phosphate group of UDP; and H158 and E331 form hydrogen bonds with the 6-OH and 3-OH groups of glucose, respectively (Li et al., 2020). In the SPP (right panel), the residues Q107, K116, and N189 binds to S6P by hydrogen bonds in the glucose ring (Fieulaine et al., 2005). Figures were prepared with ChimeraX (Pettersen et al., 2021).

(EC 3.1.3.24) is a member of the L-2-haloacid dehalogenase (HAD) superfamily of aspartate-nucleophile hydrolases, belonging to the subfamily IIB that includes SPP from plants and cyanobacteria (Albi et al., 2016). SPP carries out the second step in sucrose synthesis by removing the phosphate group from S6P, forming sucrose (Lunn et al., 2000; Fieulaine et al., 2005). The hydrolytic activity of SPP is specific to

S6P, showing little or no activity upon other sugar phosphates, such as F6P, which possesses a nearly identical phosphofructosyl moiety to S6P (Lunn et al., 2000). The mechanistic basis for the specificity of SPP to S6P against F6P appears to be related to the multiple active site contacts to the glucose ring, as revealed by a crystal structure of *Synechocystis* sp. PCC 6803 (Fieulaine et al., 2005).

Although members of the HAD superfamily generally have little overall sequence identity, they are characterized by three conserved motifs (I, II and III) related to the active site of the phosphohydrolase domain (PHD, Fieulaine et al., 2005; Figures 2A, B). All three motifs are highly conserved in SPP proteins among plants, algae, cyanobacteria, and mosses. Structurally, SPP proteins resemble a pair of "tongs" with a 'core' domain and a 'cap' domain connected by two flexible loop regions that act analogously to hinges between a closed (sucrose bound) and open (no ligand) enzyme state (PHD, Fieulaine et al., 2005). The conserved motifs that contribute to substrate binding line the interface between the two protein domains. Motif I, DXDX[T/V][L/V/I] (Figures 2A,B), is the most widely conserved among SPP sequences, and the first Asp is the functional nucleophile, which in the HAD phosphatase is transiently phosphorylated during the catalytic reaction (Collet et al., 1998; Fieulaine et al., 2005). The second Asp located in this motif is implicated in the acid–base catalysis reaction (Collet et al., 1998). Motif II, $[S/T]X_2$, contains a Ser or Thr that is generally neighbored by hydrophobic residues, and functions to bind a phosphoryl oxygen in the substrate, orienting it in the correct position for nucleophilic attack by the first Asp in motif I (Wang et al., 2001). Motif III, $KX_{18\text{-}30}[G/S][D/S]X_3[D/N]$ (Figures 2A,B), includes a conserved Lys that stabilizes the phosphorylated Asp intermediate state (Figure 2C, right panel). In addition, the two conserved Asp residues in this motif might form a system to direct water for the hydrolysis of the acyl-phosphate intermediate (Aravind et al., 1998). In a recent publication, the first Asp residue located in the motif I of the SPP domain of the bidomainal SPS_{7942} was mutated (D473A; Liang et al., 2020). This substitution inhibited sucrose synthesis, specifically the dephosphorylation of S6P to release sucrose, indicating that the Asp at position 473 is necessary for the SPP activity of the bifunctional SPS from this cyanobacterium.

In plants and several cyanobacterial species, the synthesis of sucrose is performed by a bidomainal SPS which encodes fused SPS and SPP domains on the same polypeptide (Curatti et al., 1998; Salerno and Curatti, 2003; Martinez-Noël et al., 2013; Kolman et al., 2015; Li et al., 2020; Supplementary Table S1). This is in contrast to many other cyanobacterial species where SPS and SPP are not fused, and are encoded by separate genes (Porchia and Salerno, 1996; Cumino et al., 2002; Lunn, 2002). In other words, two different domain arrangements have been described for cyanobacterial SPSs: (i) the minimal SPS unit with only a glucosyltransferase domain (GTD), found in filamentous cyanobacteria such as *Nostoc* sp. PCC 7119 (Porchia and Salerno, 1996), *Nostoc* sp. PCC 7120 (Cumino et al., 2002), and several species of unicellular cyanobacteria belonging to the genus *Gloeobacter*, *Thermosynechococcus*, and *Acaryochloris* (Blank, 2013); and (ii) the two-domain SPS prototype with both a GTD and a PHD, found in unicellular cyanobacteria such as *Synechocystis* sp. PCC 6803 (Curatti et al., 1998; Lunn et al., 1999), *Synechococcus* sp. PCC 7002 (Cumino et al., 2010) and *S. elongatus* PCC 7942 (Martinez-Noël et al., 2013; Figure 2A; Supplementary Table S1).

SPS was first reported in cyanobacteria based on characterization of a single functional GTD encoded in the filamentous cyanobacterium *Nostoc* sp. PCC 7119 (Porchia and Salerno, 1996). In this strain, two different isoforms of SPS can be found, SPS-I and SPS-II, both with similar molecular masses. The main difference between these two isoforms is their substrate specificity: whereas SPS-I has preference for UDP-Gluc, GDP-Gluc, and TDP-Gluc as substrates; SPS-II only uses UDP-Gluc and ADP-Gluc. It was previously accepted that unidomainal SPS enzymes were restricted to filamentous cyanobacterial species (Salerno and Curatti, 2003), but an extensive BLAST search in cyanobacterial genomes revealed that unidomainal SPSs are widespread in cyanobacteria, being present in species of *Gloeobacter*, *Thermosynechococcus*, *Acaryochloris*, a number of Nostocales, and other filamentous and unicellular cyanobacteria (Blank, 2013; Supplementary Table S1).

The first identification and characterization of cyanobacterial bidomainal SPS was reported in the unicellular cyanobacterium *Synechocystis* sp. PCC 6803 (Curatti et al., 1998). In independent research, SPS with fused GTD-PHD was also found in *Synechococcus* sp. PCC 7002 (Cumino et al., 2010). In addition to these species, bidomainal SPSs have been found in at least two filamentous species (*Nostoc punctiforme*, *Nodularia spumigena* CCY9414), and several unicellular cyanobacteria (e.g., *S. elongatus* PCC 7942, *Cyanobium* sp. PCC 7001, *T. elongatus* BP-1, several *Prochlorococcus* spp., and several *Synechococcus* spp.; Lunn, 2002; Martinez-Noël et al., 2013; Supplementary Table S1). Apart from a bidomainal SPS, *Synechocystis* sp. PCC 6803 has a separately encoded SPP enzyme (Lunn, 2002). The PHD of the bidomainal SPS lacks several of the conserved residues involved in the SPP function, including the critical Asp in motif I that is predicted to form an acyl-phosphate intermediate during the phosphatase reaction (Lunn, 2002; Fieulaine et al., 2005; Figure 2B). Other residues seem to be mutated in the PHD motif III of the bidomainal SPSs (indicated with an asterisk in Figure 2B), although their direct functions in SPP activity have not been described in the literature. The function of the PHD domain in biodomainal SPS is unknown, indeed some have shown these domains lack SPP activity, but it has been proposed that it might be involved in binding to newly synthesized S6P and transferring this molecule from the active site of SPS to the active site of the separately-encoded SPP in a form of metabolite channeling (Fieulaine et al., 2005). The presence of SPPs lacking enzymatic activity has also been reported in plants, where it has been suggested these non-functional protein play additional functions different from their canonical catalytic activity, for example as regulators (Albi et al., 2016).

2.1.3. SPP-like proteins

As previously discussed, SPP belongs to the class IIB subfamily of the HAD superfamily. Analysis of several cyanobacterial genomes revealed the existence of genes encoding homologous proteins of SPP (including some species encoding two or more distinct copies), but which have not been classified as SPP due to key distinctions in conserved domains. These SPP-like proteins are frequently annotated as (putative) HAD-superfamily hydrolases subfamily IIB, and while they contain the three motifs that define SPP (see above) they possess mutations in conserved residues in these motifs that distinguish them from classically defined SPPs (indicated with an asterisk in Figures 3A,B). For instance, SPP-like proteins present a conserved Gly in the fourth residue of motif I (Figure 3B) which is normally poorly conserved among cyanobacterial SPPs (Figure 2B). Conversely, the second residue of motif I is highly conserved as Leu in SPP (Figure 2B) but shows no clear conservation in SPP-like proteins (Figure 3B). Motif II seems to be conserved between SPP and SPP-like proteins, with the exception of the strictly-conserved Tyr residue in SPP which is not

conserved in SPP-like proteins (Figures 2B, 3B). In motif III, all cyanobacterial SPPs have SGN as a X_3 sequence at the end of this motif (Figure 2B), but SPP-like proteins do not maintain this sequence (Figure 3B). The first and third residues are conserved in most of the cases, but the second residue is not, with Gly substituted by Pro. Other important SPP residues include a Gln and a Lys located between motif II and III, and an Asn located in motif III, all of which are reported to participate in binding to the glucose ring of S6P *via* hydrogen bonds (Fieulaine et al., 2005; indicated by asterisks in Figure 3C).

Phylogenetic analysis of the SPP-like proteins encoded by cyanobacterial genomes revealed the existence of three main subclasses: (i) the first includes the SPP-like protein of *Synechocystis* sp. PCC 6803 and one of the two SPP-like protein paralogs encoded in the genome of members of the order Nostocales (among others); (ii) the second

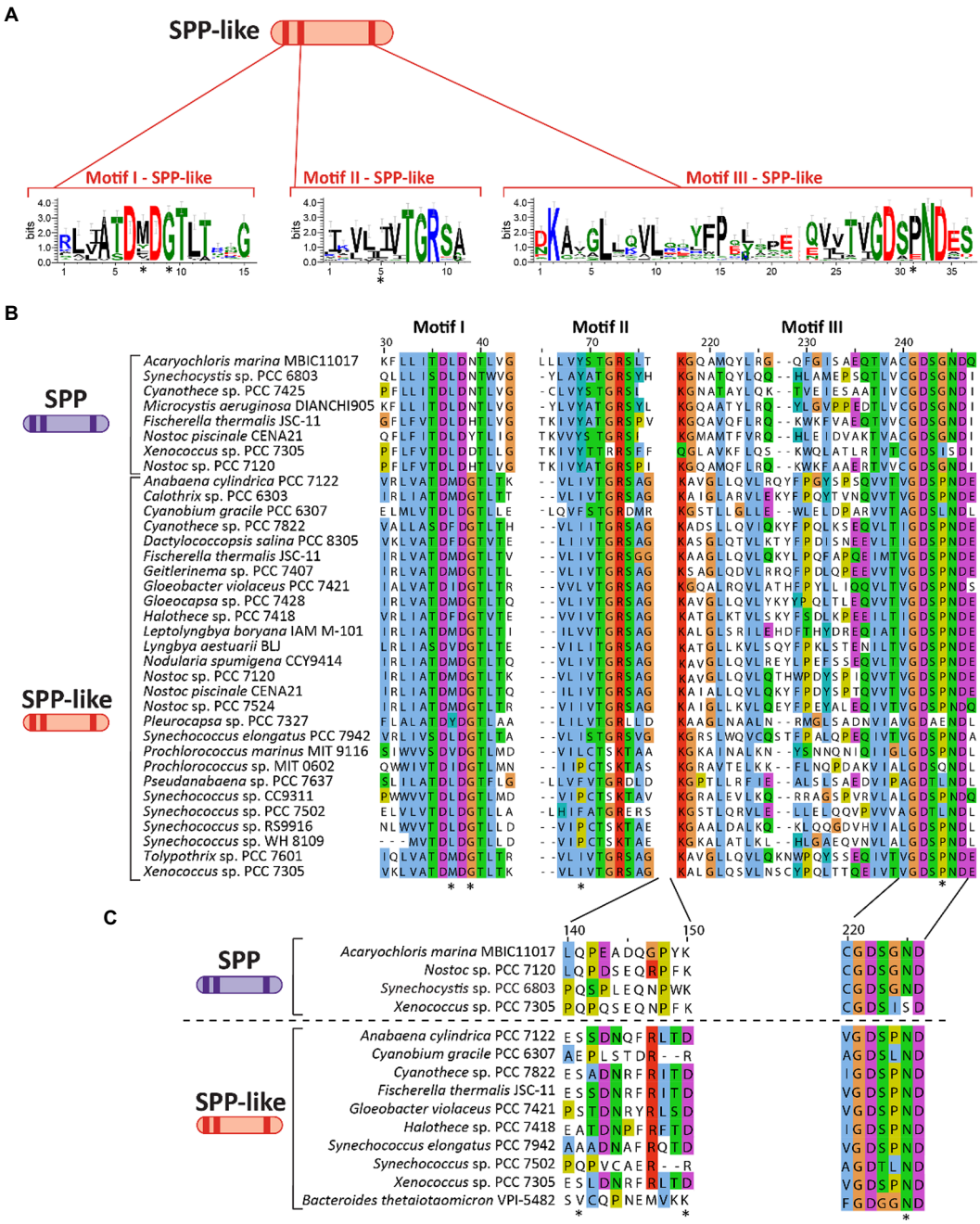

FIGURE 3
Primary structure and motifs of SPP-like proteins in cyanobacteria. **(A)** Schematic cartoon representing the motifs present in SPP-like proteins. Logos for these conserved motifs were generated using WebLogo server (Crooks et al., 2004). **(B)** Multiple sequence alignment analysis of amino acid sequences for three motifs found in SPP-like proteins, with asterisks denoting residues mutated in comparison with conserved sequence of cyanobacterial SPPs. **(C)** Multiple sequence alignment analysis of a region of amino acid residues between motif II and motif III previously implicated in S6P binding to SPP proteins. Asterisks indicate residues directly binding to the glucose ring of the S6P molecule by hydrogen bonds in a crystal structure reference (Li et al., 2020). The alignments in **(B,C)** were performed using MEGA X (Kumar et al., 2018) and visualized with the Jalview multiple sequence alignment editor using the color scheme from ClustalX (Waterhouse et al., 2009).

includes the SPP-like protein of *S. elongatus* PCC 7942 and the second of the two SPP-like protein paralogs encoded in the genome of members of the order Nostocales, and; (iii) SPP-like proteins that dominate in marine cyanobacteria, such as *Prochlorococcus* spp. (Figure 4; Supplementary Figure S1). In contrast to the conserved SPP-like subclasses, plant enzymes with verified SPP activity cluster closely to cyanobacterial classic SPPs (Figure 4; Supplementary Figure S1). For instance, the genome of *Arabidopsis thaliana* codifies four SPP isoforms, three of which exhibit SPP activity of varying catalysis rates, while the fourth one has no detectable activity (Albi et al., 2016). It has been suggested that the presence of SPP members with low/inactive catalytic activity might have regulatory functions instead (Albi et al., 2016), as it has been proposed for one of these SPP isoforms in sorghum seed germination (Jiang et al., 2015).

While canonical SPP proteins have been well reported in the literature, there exists much less direct evidence regarding the function of the SPP-like proteins. Only one member of the HAD subfamily IIB SPP-like proteins has been described in the literature, the enzyme BT4131 from the strictly anaerobic protobacteria, *Bacteroides thetaiotaomicron* VPI-5482 (Lu et al., 2005). As in the case of cyanobacterial SPP-like proteins, BT4131 is distantly related to SPP based on the phylogenetic analysis (Figure 4; Supplementary Figure S1). Substrate docking and biochemical experiments showed that BT4131 exhibited enzymatic activity on S6P and trehalose 6-phosphate, albeit with poor affinity and low rates of catalysis (Lu et al., 2005). Instead, BT4131 showed higher catalytic activity on cyclic hexose 6-phosphates and pentose 5-phosphates. To date, the function of these SPP-like proteins in cyanobacteria is unknown. One speculative possibility is that cyanobacterial SPP-like proteins play regulatory roles akin to those proposed for some plant homologs. However, it is equally possible that the aforementioned residue changes may influence or abolish the catalytic activity on S6P in these SPP-like proteins, or perhaps change their substrates entirely.

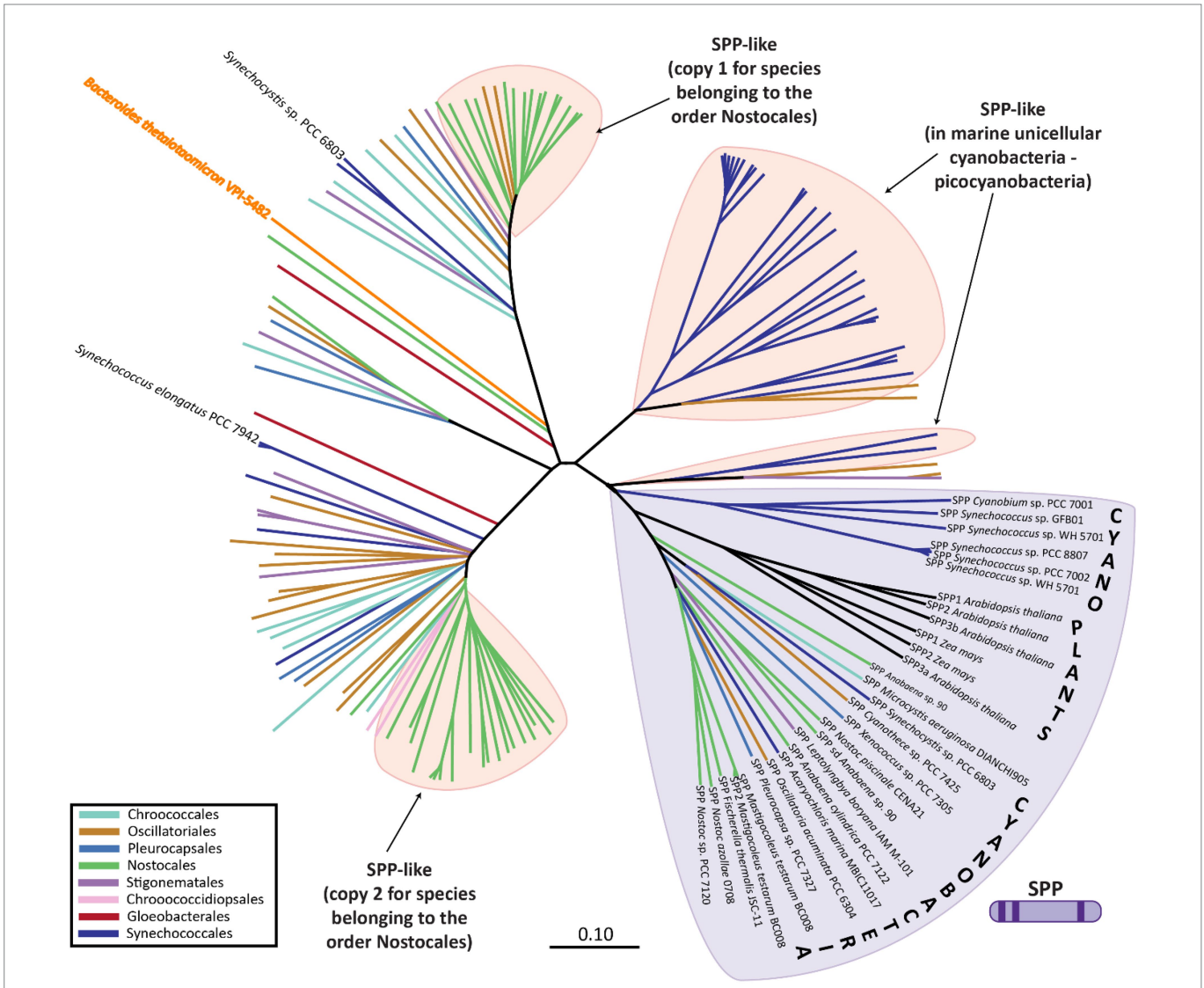

FIGURE 4
Phylogenetic analysis of SPP-like proteins encoded within different cyanobacterial species. SPP proteins are indicated by SPP followed by the name of the strain, whereas SPP-like are indicated only by the name of the strain. Unrooted neighbor-joining phylogenetic trees were constructed after sequence alignment of the SPP and SPP-like proteins using ClustalX with a BLOSSUM matrix and a bootstrap trial of 1,000. The graphical representation of the tree was generated using FigTree. Sequences were obtained from the non-redundant protein databases of the National Center for Biotechnology Information by BLAST searches. An extended version of this phylogenetic tree including all species names is available as Supplementary Figure S1.

2.2. Sucrose degradation pathways

Sucrose is a compatible solute that is transiently synthesized and accumulated during periods of the salt-stress response across many cyanobacterial species. Catabolism of sucrose is therefore required to recycle the compatible solutes after relaxation of salt stress to avoid a net loss of carbon and energy (Baran et al., 2017). Three enzymes involved in sucrose breakdown have been identified in cyanobacteria: (i) invertase (Inv), which hydrolyzes sucrose directly into glucose and fructose; (ii) amylosucrase (AMS), which catalyzes the conversion of sucrose into fructose and glucose that is often transferred to maltooligosaccharides; and (iii) sucrose synthase (SuS), which uses (A/U)DP to reversibly split sucrose into (A/U)DP-Gluc and fructose (Liang et al., 2020; Figure 1).

2.2.1. Invertases

The most broadly encoded pathway for sucrose degradation in cyanobacteria involves the enzyme Inv (EC 3.2.1.26), which irreversibly hydrolyzes sucrose into the monosaccharides, glucose and fructose (Figure 1; Supplementary Table S1). Phylogenetic analyses of Inv amino acid sequence data suggest that these enzymes originated from an ancestral Inv and their genes were transferred from cyanobacteria to plants, similarly to genes of other enzymes involved in sucrose metabolism (i.e., SPS and SPP, Vargas and Salerno, 2010). Invs are a large and diverse group of sucrose-cleaving enzymes, which can be classified partially based on their pH optimum: (i) acid Invs (Ac-Invs; β-fructofuranosidases) that possess a pH optimum in range from 4.5 to 5, and; (ii) alkaline/neutral Invs (A/N-Invs) which have a more-neutral pH optimum from 6.5 to 8 (Kirsch et al., 2019). A/N-Invs are not considered general β-fructofuranosidases since they are highly specific in catalyzing the cleavage of the α-1,2-glycosidic linkage of sucrose (Vargas et al., 2003; Vargas and Salerno, 2010; Xie et al., 2016; Liang et al., 2020). By contrast, Ac-Invs can cleave sucrose and other β-fructose-containing oligosaccharides such as raffinose and stachyose (Sturm, 1999). Bioinformatic analyses have shown that cyanobacterial genomes only encode A/N-Invs, but not Ac-Invs (Xie et al., 2016; Wan et al., 2018), whereas Ac-Invs can be mainly found in heterotrophic bacteria, yeasts, and plants (Tauzin and Giardina, 2014; Nadeem et al., 2015). In plants, A/N-Invs can be found in the cytosol, mitochondria, and/or in plastids, whereas Ac-Invs are frequently localized to the vacuolar space or bound to the cell wall (Tauzin and Giardina, 2014).

Early reports describing the hydrolysis of sucrose by A/N-Inv activity were published in *Trichormus variabilis* (Schilling and Ehrnsperger, 1985) and in *Scytonema* spp. (Page-Sharp et al., 1999). In addition, the first isolation and characterization of a cyanobacterial Inv was made in *Nostoc* sp. PCC 7120 (Vargas et al., 2003), which possesses two A/N-Inv encoded by *invA* and *invB* genes (Supplementary Table S1; Page-Sharp et al., 1999). By insertional inactivation, it has been demonstrated that InvA has a regulatory role controlling carbon flux from vegetative cells to heterocysts (López-Igual et al., 2010). The absence of this enzyme within the vegetative cells affects heterocyst differentiation due to a C/N imbalance in the filament, although it has been also speculated that sucrose or a product of its degradation might be regulating this process (Cumino et al., 2007; Vargas et al., 2011; Ehira et al., 2014). In addition, InvB activity is exclusively related to heterocysts, where it has an important function in heterocyst development, nitrogen fixation, and diazotrophic growth (López-Igual et al., 2010; Vargas et al., 2011; Xie et al., 2018). In a recent report, it was demonstrated that in *Synechocystis* sp. PCC 6803, the only enzyme responsible for *in vivo* sucrose degradation is an Inv (Kirsch et al., 2018).

2.2.2. Amylosucrases

AMS (EC 2.4.1.4) is a glucosyltransferase that catalyzes the hydrolysis of the glycosidic bond in sucrose, leading to the release of glucose and fructose. Then, the released glucose is used to form α-1,4-linked linear insoluble glucans (amylose-like polymers; Potocki de Montalk et al., 2000; Figure 1). AMS belongs to glycoside hydrolase (GH) family 13 (the α-amylase family), and is organized in five domains: N, A, B, C (common to all proteins to the GH family 13), and an additional special domain called the B′ domain (only found in AMS; Skov et al., 2001). The A, B, and B′ domains form of an active site pocket, directly related to the activity of AMS (Skov et al., 2001). The first reported AMS in a cyanobacterium was in *Synechococcus* sp. PCC 7002, where sucrose synthesis genes (*sps* and *spp*) are grouped in the same transcriptional unit with fructokinase and AMS encoding genes (Perez-Cenci and Salerno, 2014). In comparison to Inv, it is relatively rare for cyanobacterial species to encode AMS, and the presence of this gene in the genome is frequently associated with the absence of other proteins able to breakdown sucrose (Supplementary Table S1).

2.2.3. Sucrose synthases

SuS (EC 2.4.1.13) is a glucosyltransferase that can catalyze both the synthesis and cleavage of sucrose, but appears to be active principally in the cleavage reaction *in vivo* (Porchia et al., 1999; Curatti et al., 2002; Figure 1). The reversible cleavage of sucrose yields fructose and ADP-Gluc. SuS activity was first reported in cyanobacteria in *T. variabilis* ATCC 29413 (Schilling and Ehrnsperger, 1985) and *Nostoc* sp. PCC 7119 (Porchia et al., 1999). SuS has been mainly found in heterocyst-forming strains, where it seems to play an essential role in the control of carbon fluxes originating form vegetative cells through the cleavage of sucrose in the heterocysts (Porchia et al., 1999; Curatti et al., 2000, 2002, 2006, 2008). Outside of its roles in localizing the breakdown of sucrose in filamentous cyanobacteria, SuS is also reported in several unicellular cyanobacterial strains (Kolman et al., 2012; Figueroa et al., 2013; Tanabe et al., 2019; Supplementary Table S1). SuS has been implicated in the regulation of glycogen synthesis through its capacity to provide sugar nucleotide substrates (i.e., ADP-Gluc) required for elongation of α-1,4-glucoside chains (Cumino et al., 2002, 2007), a process that seems to be controlled by nutritional and environmental signals (Curatti et al., 2008).

Phylogenetic analysis suggests that a gene duplication of the GTD from a SPS-like gene and an addition of a N-terminal extension gave rise to SuS in most cyanobacteria (Cumino et al., 2002). These events took place before the branching of filamentous heterocyst-forming cyanobacteria. The occurrence of SuS in most-recently radiated cyanobacterial species, such as *Gloebacter violaceus* PCC 7421 and *Microcystis aeruginosa* NIES-4325, might be due to a more recent lateral gene transfer event (Blank, 2013; Tanabe et al., 2018, 2019).

3. Roles of sucrose in cyanobacteria

3.1. Sucrose as compatible solute

Cyanobacteria are ubiquitous organisms distributed widely across habitats and including terrestrial, aquatic, hypersaline waters, salt pans, and extreme environments such as deserts and hot thermal vents

(Whitton and Potts, 2002). Cyanobacteria have evolved specific mechanisms to cope with the associated stress conditions of these ecosystems. In aquatic environments, salinity fluctuations are very common due to changes in freshwater inflow by climate, weather, and diurnal tidal currents. High salt concentrations promote loss in cytosolic water availability and increased ion concentrations that be destabilize many biomolecules (Klähn and Hagemann, 2011; Kolman et al., 2015; Kirsch et al., 2019). Cyanobacteria utilize the "salt-out strategy" for osmotic acclimation of the cytoplasm to changing salt concentrations (Pade and Hagemann, 2014). Briefly, the accumulation of small organic molecules called compatible solutes (including sucrose) acts to combat the loss of cytoplasmic water and corresponding drop in turgor pressure that normally accompanies a high extracellular osmotic pressure. In tandem, cyanobacteria engage numerous transporters that act to pump out the continuous influx of inorganic ions (e.g., Na^+ and Cl^-) that pass through the cell membrane under conditions of high external ionic pressure (Keshari et al., 2019).

Compatible solutes are organic molecules with low molecular masses without a net charge, which can accumulate to high (molar) concentrations in the cytoplasm without interfering with the cellular metabolism (Klähn and Hagemann, 2011). In cyanobacteria, different compatible solutes have been described and can be classified in the following substance classes: sugars (trehalose, sucrose), heterosides [glucosylglycerol (GG), glucosylglycerate (GGA)], amino acid derivatives (glycine betaine, glutamate betaine, homoserine betaine), polyols (glycerol), amino acids (proline), and organosulfurs (dimethylsulfoniopropionate; Kirsch et al., 2019; Kageyama and Waditee-Sirisattha, 2022). A correlation has been established between the class of the dominant compatible solute used by a given cyanobacterial species and its degree of exposure to salt stress within its natural habitat (Reed et al., 1984; Reed and Stewart, 1985). In general, freshwater strains with low halotolerance usually accumulate the disaccharides sucrose and/or trehalose as a compatible solute. Marine cyanobacteria accumulate the heterosides GG and GGA as osmolytes and are able to tolerate moderate salt concentrations. Finally, glycine betaine and glutamate betaine are preferentially synthesized as compatible solutes in halophilic species that inhabit extremely saline environments (Mackay et al., 1984; Hagemann, 2011). However, there are few exceptions to this classification. Notably, the widespread marine picoplanktonic *Prochlorococcus* strains appear to utilize sucrose as the preferred compatible solute (Klähn et al., 2010). In addition, some *Prochlorococcus* and *Synechococcus* strains also synthesize glycine betaine as well as GGA (Klähn et al., 2010).

3.2. Control of sucrose synthesis and degradation enzymes by ions

Commonly, the activity of enzymes responsible for synthesis and degradation of compatible solutes are regulated directly by allosteric binding of specific ions (Page-Sharp et al., 1999; Marin et al., 2002; Kirsch et al., 2019; Liang et al., 2020). In *S. elongatus* PCC 7942, SPS_{7942} activity is regulated by inorganic ions, Na^+ and Cl^-, which activate the SPS domain of this bifunctional protein, but have relatively little impact on the enzymatic activity of the SPP domain (Liang et al., 2020). The same ion-induced SPS activation has also been observed in the closely related strain, *S. elongatus* PCC 6301 (Hagemann and Marin, 1999), and other unicellular strains, such as *Synechocystis* sp. PCC 6803 (Hagemann and Marin, 1999; Desplats et al., 2005) and *M. aeruginosa* PCC 7806 (Kolman and Salerno, 2016). Inv activity in *S. elongatus* PCC 7942 is also regulated in an ion-dependent manner, showing decreased catalysis under elevated ion concentrations. The inhibition of Inv by ions combines with ion-induced SPS activation, promoting overall intracellular sucrose accumulation (Liang et al., 2020). The same regulation of invertase has been described for *Synechocystis* sp. PCC 6803 (Kirsch et al., 2018), suggesting that this is a fairly widespread mechanism that contributes to sucrose accumulation under salt stress.

Gene expression of sucrose synthesis enzymes is also controlled in a salt-induced manner (Cumino et al., 2010; Perez-Cenci and Salerno, 2014; Kolman and Salerno, 2016). For example, *sps* gene expression is usually transcriptionally activated upon salt addition, promoting sucrose synthesis proportional to the severity of osmotic stress (Kirsch et al., 2019). In *Synechococcus* sp. PCC 7002, a salt treatment increased the transcript levels of *sps* and *spp*, genes that organized together in an operon with AMS and fructokinase (Cumino et al., 2010; Perez-Cenci and Salerno, 2014). Similarly, *M. aeruginosa* PCC 7806 also contains a sucrose gene cluster including *spsA*, *susA*, and *sppA* that are all stimulated by salt (Kolman and Salerno, 2016). The transcript levels of *susA* were also increased in *M. aeruginosa* PCC 7806 and *G. violaceus* PCC 7421 cells after a salt treatment (Kolman et al., 2012). More recently, it was shown that transcription of the *sps* gene is upregulated after the addition of NaCl to *S. elongatus* PCC 7942 (Liang et al., 2020). In *Synechocystis* sp. PCC 6803, some sensors have been identified to be possibly related with perceiving and transducing signals of salt and hyperosmotic stresses (Marin et al., 2002; Shoumskaya et al., 2005; Liang et al., 2011), and a two-component response regulator was confirmed to control sucrose synthesis in *Nostoc* sp. PCC 7120 (Ehira et al., 2014). Finally, in some species, NaCl treatment has been shown to directly increase SPS specific activity and concurrently activate *sps* gene expression (Hagemann and Marin, 1999; Salerno et al., 2004).

3.3. Other functions of sucrose in cyanobacteria

Apart from its role as a compatible solute, sucrose acts in other cellular pathways. Sucrose and trehalose are considered major compatible solutes that enhance drought tolerance in cyanobacteria (Rajeev et al., 2013; Wang et al., 2018; Lin P. C. et al., 2020; Khani-Juyabad et al., 2022), though their functions in desiccation tolerance are less rigorously characterized. Sucrose also has well-established roles as a fixed carbon carrier molecule in some filamentous species, where it is produced in vegetative cells and catabolized in the heterocysts of nitrogen-fixing cyanobacterial species (Nürnberg et al., 2015). Sucrose acts a molecule to carry carbon and energy equivalents from vegetive cells to heterocysts, where it is consumed in part to drive the ATP-and NADH-requiring nitrogenase reactions (Juttner, 1983; Cumino et al., 2007; López-Igual et al., 2010; Vargas et al., 2011). It is proposed that sucrose transport primarily occurs through cell–cell septal junctions (Nürnberg et al., 2015).

Finally, it has been speculated that sucrose might also act as a signaling molecule in cyanobacteria (Desplats et al., 2005). In higher plants, sucrose metabolism is not only essential for the allocation of carbon resources but also participates in a regulatory network that coordinates metabolism and development (Curatti et al., 2006). Sucrose seems to be a versatile molecule with multiple roles in cyanobacteria, but most of them are poorly understood, raising the possibility that this sugar has underappreciated functions that remain unexplored.

4. Engineering cyanobacteria to produce sucrose

Innovations in biotechnology have taken advantage of aquatic photosynthetic organisms' ability to create valuable products (e.g., lipids, antioxidants, pigments) to cope with environmental stressors (Chen et al., 2017; Morone et al., 2019). As a bioproduct naturally synthesized at high levels by some species of cyanobacteria in response to salt stress, sucrose has garnered attention for its potential as an alternative carbohydrate feedstock for higher-value goods (Hays and Ducat, 2015; Zhang et al., 2021). Sucrose generated by cyanobacteria could offer a number of advantages relative to plant-based feedstock crops, including potentially higher photosynthetic efficiencies and reduced requirements for potable water or arable land. Here, we review recent strategies employed to make cyanobacterial bioproduction of sucrose more productive and affordable.

4.1. Engineered heterologous transporters for sugar export

As discussed above, cyanobacteria can accumulate osmoprotectants up to hundreds-of-millimolar concentrations when exposed to hypersaline conditions (e.g., sucrose, trehalose, GG; Hagemann, 2011; Klähn and Hagemann, 2011). For instance, under moderate salt stress (200 mM NaCl), the common freshwater model cyanobacterium *S. elongatus* PCC 7942 accumulates nearly 300 mM intracellular sucrose (calculated based on a culture volume basis), representing a significant portion of the cell biomass (Suzuki et al., 2010). Although this degree of metabolite production presents an industrial and agricultural opportunity, cytosolic volume constrains how much sucrose can be accumulated: the costs associated with cyanobacterial cell recovery, lysis, and processing would likely exceed the economic value of the commodity products contained in the cytosol (Prabha et al., 2022). Therefore, secreting sugars into the supernatant for collection has been proposed as a more financially viable strategy. For this purpose, cyanobacteria have been engineered to express heterologous transporters capable of exporting lactate and hexoses (Niederholtmeyer et al., 2010; Angermayr et al., 2012).

Similarly, *S. elongatus* PCC 7942 was originally engineered to export sucrose by expressing sucrose permease (*cscB*) from *Escherichia coli* ATCC 700927 (Ducat et al., 2012), and multiple cyanobacterial species have since been similarly modified by different research teams (Table 1). In its native host, CscB is a sucrose/proton symporter that typically operates by utilizing the free energy of the proton gradient to import both molecules (Vadyvaloo et al., 2006). By contrast, during periods of cyanobacterial sucrose synthesis, internal sucrose concentrations greatly exceed external levels causing reversal of chemical gradients and driving sucrose export through CscB instead. CscB-expressing, sucrose-exporting *S. elongatus* PCC 7942 strains can secrete up to 80% of photosynthetically fixed carbon as sucrose, diverting these resources away from the accumulation of cellular biomass (Ducat et al., 2012). Although efforts to scale-up cyanobacterial sucrose production have not yet come to fruition (e.g., Proterro; Aikens and Turner, 2013), it has been estimated that such cyanobacterial strains have the potential to produce comparable amounts of sugar to traditional plant-based carbohydrate feedstocks. Realizing the promise of cyanobacterial sucrose is likely to require efforts to address problems of cyanobacterial/microalgal cultivation (beyond the scope of this review, but see Su et al., 2017; Khan et al., 2018) as well as strategies to maximize bioproduction rates.

4.2. Increasing metabolic flux to sucrose pathways

Published strategies for improving rates of cyanobacterial sucrose productivity generally fall into two related strategies: increasing carbon flux towards the synthesis of sucrose through the upregulation of relevant biosynthetic activities, or by reducing the loss of carbon to competing pathways or sucrose reuptake. Perhaps the most straightforward approach for improving sucrose titers has been the overexpression of genes related to sucrose biosynthesis. Several studies have now found that flux leading to sucrose production can be most impacted by increasing the activity of SPS (Du et al., 2013; Duan et al., 2016; Lin P. C. et al., 2020), which is largely intuitive given that this enzyme catalyzes a commitment step to sucrose biosynthesis. Significant increases in sucrose production can be found in strains overexpressing SPS, even without allowing for sucrose export. First reported in *Synechocystis* sp. PCC 6803, a strain engineered to overexpress its native SPS (SPS_{6803}) accumulated nearly twice as much intracellular sucrose than its wild-type counterpart (Du et al., 2013). Likewise, when the native SPS in *S. elongatus* PCC 7942 was overexpressed, internal sucrose concentrations were 93% higher than in wild-type (Duan et al., 2016). In addition, pairing SPS overexpression with sucrose export further increases total sucrose yields. When SPS_{7942} and CscB were co-overexpressed in *S. elongatus* PCC 7942, there was a 74% increase in sucrose compared to the CscB-only strain (Table 1; Duan et al., 2016), yet the nature of the SPS homolog that is overexpressed can strongly influence the degree to which sucrose production is improved. SPS_{7942} is bidomainal and bifunctional (i.e., possessing active GTD and PHD domains), in contrast to SPS_{6803} which is also bidomainal, but has a non-functional PHD domain and is regulated distinctly from SPS_{7942} (Curatti et al., 1998; Lunn et al., 1999; Gibson et al., 2002). However, the partial-functionality of SPS_{6803} does not mean it is less effective, as heterologous co-overexpression of SPS_{6803} and CscB in *S. elongatus* PCC 7942 increases sucrose production relative to overexpression of the native SPS_{7942} (Abramson et al., 2016; Dan et al., 2022; Table 1). It is curious that SPS_{6803} is a more effective enzyme for rerouting carbon flux towards sucrose bioproduction, given that it lacks a functional SPP domain (*S. elongatus* PCC 7942 encodes other endogenous SPP proteins in the examples above), so it is possible that this observation is related either to the manner in which salt-ions can regulate the function of some SPS domains (Liang et al., 2020), or to other unknown functions for SPP and/or SPP-like domains other than S6P phosphatase activity (see 2.1.3. SPP-like proteins).

While the overexpression of SPS has yielded substantial improvements, this strategy has not been equally successful with other proteins in the sucrose biosynthetic pathway. Overexpression of SPP from *Synechocystis* sp. PCC 6803 (SPP_{6803}) either had no effect on, or decreased sucrose productivity in *S. elongatus* PCC 7942 or *Synechococcus elongatus* UTEX 2973 (Du et al., 2013; Lin P. C. et al., 2020). Similarly, overexpression of UDP-Gluc pyrophosphorylase (UGP), the protein producing UDP-Gluc as a substrate for SPS, led to less sucrose secretion (Ducat et al., 2012; Du et al., 2013). Only when these three enzymes were overexpressed simultaneously (i.e., SPS_{6803}, SPP_{6803}, and UGP), were sucrose levels increased in comparison with SPS_{6803}-only strain, albeit marginally (Du et al., 2013; Table 1).

TABLE 1 Productivity and genetic modifications of sucrose-producing cyanobacteria.

Species[a]	Overexpressed[b,*]	Downregulated[c]	Maximum productivity[d]	Salt for SPS	Sucrose promoter	Reference
*Syn*7002	–	–	24 mol 10^{-17} cells[e]	1 M NaCl	–	Xu et al. (2013)
	–	*glgA-I, glgA-II*	71 mol 10^{-17} cells[e]	1 M NaCl	–	
*Syn*6803	*cscB*, sps_{6803}, spp_{6803}, *ugp*	*ggpS*, *ggtCD*	0.69 mg L^{-1} h^{-1}[f]	400 mM NaCl	P_{petE}	Du et al. (2013)
	sps_{6803}, spp_{6803}, *ugp*	*ggpS*	3.1 mg L^{-1} h^{-1}[f]	600 mM NaCl	P_{petE}	
*Syn*7942	*cscB*	–	3.6 mg gDW^{-1} h^{-1}	200 mM NaCl	–	Duan et al. (2016)
	cscB, sps_{7942}	–	6.2 mg gDW^{-1} h^{-1}	200 mM NaCl	P_{trc}	
*Syn*7942	–	Synpcc7942_1125	5.9 mg L^{-1} OD_{730}^{-1} h^{-1}[e]	150 mM NaCl	–	Qiao et al. (2019)
	–	*manR* (Synpcc7942_1404)	6.7 mg L^{-1} OD_{730}^{-1} h^{-1}[e]	150 mM NaCl	–	
*Syn*7942	*cscB*, sps_{6803}	–	5.6 mg L^{-1} h^{-1}	–	P_{cpcB}	Dan et al. (2022)
	cscB, sps_{6803}, *glgP*	–	6.9 mg L^{-1} h^{-1}	–	P_{cpcB}	
*Syn*6803	*cscB*, sps_{6803}	*ggpS*	6.3 mg L^{-1} h^{-1}	400 mM NaCl	P_{trc}	Thiel et al. (2019)
*Syn*7942	*cscB*, sps_{7942}, *glgC*	–	8 mg L^{-1} h^{-1}	150 mM NaCl	P_{trc}	Qiao et al. (2018)
*Syn*6803	*cscB*, sps_{6803}, spp_{6803}, *ugp*	*invA* (sll0626), *ggpS*, *ggtCD*	10.1 mg L^{-1} h^{-1}	200 mM NaCl	P_{petE}	Kirsch et al. (2018)
*Syn*7942	*cscB*	–	10.4 mg L^{-1} h^{-1}	150 mM NaCl	–	Löwe et al. (2017)
*Syn*7942	*cscB*	–	11 mg L^{-1} h^{-1}	150 mM NaCl	–	Li C. et al. (2022)
*Syn*7942	*cscB*	–	16.7 mg L^{-1} h^{-1}	106 mM NaCl	–	Hays et al. (2017)
*Syn*7942	*cscB*, sps_{6803}	–	30 mg L^{-1} h^{-1}	–	P_{trc}	Abramson et al. (2016)
*Syn*2973	*cscB*	–	24.6 mg L^{-1} h^{-1}	150 mM KCl	–	Song et al. (2016)
			35.5 mg L^{-1} h^{-1}	150 mM NaCl	–	
*Syn*7942	*cscB*	–	28 mg L^{-1} h^{-1}	150 mM NaCl	–	Ducat et al. (2012)
	cscB	*invA*, *glgC*	36.1 mg L^{-1} h^{-1}	150 mM NaCl	–	
*Syn*7942	*cscB*, sps_{6803}, *rpaB*	–	48 mg L^{-1} h^{-1}	–	P_{trc}	Abramson et al. (2018)
*Syn*2973	*cscB*, sps_{6803}, spp_{6803}	–	22.2 mg L^{-1} h^{-1}	–	P_{trc1O}, induced	Lin P. C. et al. (2020)
			47.2 mg L^{-1} h^{-1}	–	P_{trc1O}, uninduced	
	cscB	–	79.2 mg L^{-1} h^{-1}	150 mM NaCl	–	

[a]*Syn7002, Synechococcus* sp. PCC 7002; *Syn7942, Synechococcus elongatus* PCC 7942; *Syn2973, Synechococcus elongatus* UTEX 2973; *Syn6803, Synechocystis* sp. PCC 6803.

[b]*cscB*, sucrose permease; *glgC*, ADP-glucose pyrophosphorylase; *glgP*, glycogen phosphorylase; *rpaB*, regulator of phycobilisome-associated B; *spp*, sucrose phosphate phosphatase; *sps*, sucrose phosphate synthase; *ugp*, UDP-glucose pyrophosphorylase.

[c]Genes are down-regulated or knocked out; *ggpS*, glucosylglycerol (GG)-phosphate synthase; *ggtCD*, GG transport system permease; *glgA-I/glgA-II*, glycogen synthase; *glgC*, ADP-glucose pyrophosphorylase; *invA*, invertase; *manR* (Synpcc7942_1404), manganese sensing response regulator; Synpcc7942_1125, histidine-containing phosphotransfer.

[d]Approximated extracellular sucrose values provided or calculated from titers.

[e]Intracellular sucrose yields.

[f]Intracellular and extracellular (total) sucrose yields.

*Subscript in *sps* and *spp* indicates the strain that it comes from (i.e., 6803 for *Synechocystis* sp. PCC 6803, 7942 for *S. elongatus* PCC 7942).

Another successful approach for improving the flux of carbon to sucrose biosynthesis is to accelerate the rate of the upstream carbon supply from the CBB. Multiple strains of cyanobacteria have been engineered to secrete sucrose through the heterologous expression of *cscB*, but the highest yields to-date have been obtained from strains with a more rapid metabolism and higher light tolerance relative to classic laboratory models (Table 1). *S. elongatus* UTEX 2973 is a recently re-characterized species that is 99.99% identical to *S. elongatus* PCC 7942, but has a doubling time as fast as ~2 h (compared to ~5–9 h for *S. elongatus* PCC 7942), and is more tolerant of high-light and high-temperature conditions (Kratz and Myers, 1955; Yu et al., 2015; Adomako et al., 2022). Expression of *cscB* in *S. elongatus* UTEX 2973 led to the development of strains with relatively high sucrose productivities (Song et al., 2016; Lin P. C. et al., 2020). A high sucrose titer was originally reported in such strains when exposed to 150 mM NaCl, reaching approximately 80 mg L^{-1} (Song et al., 2016). Lin et al. also created a *S. elongatus* UTEX 2973-*cscB* strain, and observed an even greater sucrose titer at 8 g L^{-1} at 150 mM NaCl, averaging out to 1.9 g L^{-1} day^{-1}, over 2-fold higher than the productivities of *S. elongatus* PCC 7942, representing the highest sucrose titer published thus far (Lin P. C. et al., 2020), and illustrating the potential benefits of utilizing fast-growing strains that can reach higher densities.

Somewhat surprisingly, activation of the sucrose export pathway itself has been reported to increase the overall photosynthetic flux in some cyanobacterial strains. In *S. elongatus* PCC 7942, when sucrose synthesis pathways are placed under inducible promoters, a variety of enhancements in features related to photosynthesis have been reported in the hours following activation of the pathway (Ducat et al., 2012;

Abramson et al., 2016; Santos-Merino et al., 2021b; Singh et al., 2022). The quantum efficiency of photosystem II, rate of oxygen evolution, relative rate of electron flux through the photosynthetic electron transport chain, oxidation status of photosystem, and rate of carbon fixation are all increased (Ducat et al., 2012; Abramson et al., 2016; Santos-Merino et al., 2021b). The latter observation is correlated with an increase in Rubisco abundance that was revealed by proteomic analysis >24 h following induction of sucrose export, and a concomitant increase in carboxysome number (Singh et al., 2022). While the mechanisms underlying these changes in photosynthetic performance are not well understood, it has been hypothesized that they arise from a relaxation in "sink limitations" on photosynthesis that can arise when the downstream consumption of products of photosynthesis (e.g., ATP, NADPH, CBB outputs) is insufficient to keep up with the supply (Santos-Merino et al., 2021a). Stated differently, when carbon fixation is not the rate-limiting step of cell metabolism (e.g., under enriched CO_2 atmospheres commonly used in laboratory conditions), the expression of a heterologous pathway may act as an additional "sink" and bypass downstream limitations of cell growth and division. While this remains a speculative possibility, the relaxation of acceptor-side limitations on photosystem I suggests that sucrose secretion pathways (or other heterologous metabolic sinks) may utilize "excess" light energy that might otherwise be lost to photosynthetic inefficiencies under certain conditions (Abramson et al., 2016; Santos-Merino et al., 2021b). Uncovering the mechanisms underlying the photosynthetic phenotypes coupled to sucrose export might allow even greater improvements in photosynthesis and/or sucrose bioproduction.

4.3. Reducing metabolic flux to competing pathways

The alternative strategy to boost sucrose production is to improve the pool of sucrose or sucrose precursors by reducing flux to pathways that compete with sucrose biosynthesis for either substrates or total carbon pools. A straightforward example is to eliminate the dominant route for sucrose breakdown, such as the Inv proteins that are a dominant route of sucrose hydrolysis in many cyanobacterial models. In a recent report, inactivation of the *Synechocystis* sp. PCC 6803 invertase increased accumulated sucrose by 10-fold in both salt and salt-free conditions (Kirsch et al., 2018). These results were of higher magnitude, but similar trajectory to reports in other cyanobacteria, such as in sucrose-exporting *S. elongatus* PCC 7942 where a Δ*invA* background exhibited a 15% increase in extracellular sucrose (Ducat et al., 2012).

Glycogen is a storage molecule of cyanobacteria that is a significant alternative carbon sink, yet inhibiting glycogen synthesis has yielded variable results on sucrose secretion. For example, knockout of the two glycogen synthase genes (*glgA-I* and *glgA-II*) of *Synechococcus* sp. PCC 7002 led to an accumulation of three times more sucrose than wild-type under hypersaline conditions (Xu et al., 2013; Table 1). However, when another glycogen synthesis gene, ADP-glucose pyrophosphorylase (*glgC*), was downregulated in sucrose-secreting *S. elongatus* PCC 7942, there was only a minor or insignificant increase in sucrose (Qiao et al., 2018). GlgP is responsible for hydrolyzing glycosidic bonds in glycogen to release glucose-1-phosphate, so it was theorized that increasing GlgP activity would mobilize carbon from the glycogen pool for sucrose biosynthesis. However, when GlgP was overexpressed in sucrose-secreting strains of *S. elongatus* PCC 7942 with its native SPS, there were no changes in glycogen content and a decrease in sucrose was observed (Ducat et al., 2012; Dan et al., 2022), while heterologous expression of both SPS_{6803} and GlgP overexpression reduced glycogen content while increasing sucrose secretion by 2.4-fold (Dan et al., 2022). The variability in sucrose production of glycogen-deficient strains might be related to the pleotropic cellular deficiencies of these strains, including reduced growth, reduced O_2 evolution and consumption, abnormal pigmentation, and light sensitivity (Suzuki et al., 2010; Ducat et al., 2012; Gründel et al., 2012; Xu et al., 2013; Qiao et al., 2018). These phenotypes align with a potential broader role for glycogen beyond carbon storage, which may include buffering against periods of starvation, oxygenic stress, high-light stress, salt stress, or diurnal/transient changes in light availability (Luan et al., 2019; Shinde et al., 2020). Given the increasing recognition of regulatory roles of glycogen, more nuanced strategies may be required to regulate the flux of carbon towards glycogen synthesis in order to reliably improve sucrose bioproduction (Huang et al., 2016).

In some cyanobacterial strains that utilize other compatible solutes as the dominant metabolite for osmoprotection, synthesis of these osmoprotectant compounds may compete with sucrose biosynthesis. One example is, GG, the primary solute utilized by moderately halotolerant cyanobacteria such as *Synechocystis* sp. PCC 6803 (Klähn and Hagemann, 2011). When GG-phosphate synthase (GgpS), the enzyme that generates a GG precursor, was knocked out in *Synechocystis* sp. PCC 6803, increased flux of carbon to sucrose production was reported (Du et al., 2013; Kirsch et al., 2019; Thiel et al., 2019; Table 1). A GgpS mutant incapable of generating GG under salt stress instead accumulated nearly 1.5-fold more sucrose than wild-type, although these engineered strains also exhibited growth inhibition at lower salt concentrations that would be well tolerated by wild-type lines (Du et al., 2013).

While most studies focus on restricting metabolic pathways that consume cellular carbon resources, downregulation of processes that compete for reducing equivalents may also be an alternative approach to engineering strains with high-sucrose productivity. Flavodiiron proteins are part of cyanobacterial photoprotective systems that are engaged during periods of redox stress (e.g., fluctuating light) and can direct electrons from an over-reduced photosynthetic electron transport chain to oxygen (Allahverdiyeva et al., 2015). The flavodiiron-catalyzed reaction is essentially a water–water cycle that dissipates potential energy from reducing equivalents generated from photosynthetic light reactions, but this reaction appears to be important for preventing photodamage under dynamic light conditions (Allahverdiyeva et al., 2013). Knockout of flavodiiron proteins Flv1 and Flv3 in *S. elongatus* PCC 7942 could boost sucrose production in a $cscB/sps_{6803}$ expressing background (Santos-Merino et al., 2021b). Furthermore, activation of sucrose secretion pathways could partially compensate for the loss of Flv1/Flv3 under transient light changes, further suggesting that heterologous metabolic sinks may have some limited ability to utilize "overpotential" on the photosynthetic electron transport chain (Santos-Merino et al., 2021b).

4.4. Altering regulatory networks to increase sucrose synthesis

Sucrose biosynthesis is a natural component of many cyanobacterial adaptive responses, so a deeper understanding of the regulatory networks that control this process could allow researchers to manipulate

sucrose production in the absence of abiotic stressors. In this context, a couple of studies have reported promising improvements in sucrose secretion rates by altering cyanobacterial two-component regulatory proteins, although the specific mechanisms remain uncertain. In a screen of all two-component regulatory factors in *S. elongatus* PCC 7942, Qiao and colleagues identified genes indirectly linked to sucrose productivity, glycogen accumulation, and photosynthetic activity (Qiao et al., 2019). The partial deletion of ManR, a protein that plays a regulatory role in Mn^{2+} uptake (Ogawa et al., 2002; Yamaguchi et al., 2002; Zorina et al., 2016), increased sucrose by 60%, a complete knockout of Synpcc7942_1125 increased sucrose by 41% (Qiao et al., 2019; Table 1). In a separate study, overexpression of the two-component protein regulator of phycobilisome assembly B (*rpaB*) reproduced a growth-arrest phenotype in *S. elongatus* PCC 7942 (Moronta-Barrios et al., 2013), and increased sucrose secretion in a *cscB*-expressing background (Abramson et al., 2018; Table 1). It was suggested that the growth arrest restricted carbon flux to many downstream pathways that might otherwise compete for sucrose biosynthesis, though a more specific alteration in regulatory processes controlling sucrose synthesis could not be excluded. While the number of studies is still limited, two-component signaling pathways have so far proven to be a promising strategy to improve sucrose yields, though our mechanistic understanding for these phenotypes is far from complete.

5. Applications of sucrose production in cyanobacterial co-culture

While the biotechnological focus for cyanobacteria has predominantly been upon direct synthesis of high-value products (Ducat et al., 2011; Knoot et al., 2018), there is growing interest in utilizing sugar-producing cyanobacteria for indirect bioproduction. This approach involves the use of carbohydrate-secreting cyanobacteria that support the growth of co-cultivated heterotrophic microbes. Co-cultures become "one-pot" reactions where cyanobacteria specialize in photosynthetic metabolism to supply carbon to a heterotroph, which in turn performs the metabolic labor of converting the carbon to higher-value goods or services (Hays and Ducat, 2015; Ortiz-Reyes and Anex, 2022). In this section, we will cover the modular nature of synthetic microbial consortia designed using sucrose-secreting cyanobacteria, their applications, and their future opportunities and challenges.

5.1. Potential advantages of modular microbial platforms

Microbial bioproduction is now a well-recognized approach that harnesses metabolic diversity for synthesis of valuable chemicals (e.g., polymers, fuels, pharmaceuticals) as an alternative to traditional environmentally unsustainable processes (Tsuge et al., 2016; Wendisch et al., 2016; Liu and Nielsen, 2019; Zhong, 2020; Wu et al., 2021). Multiple decades of sustained investments in microbial research, prospecting, and genetic engineering have yielded a wealth of bacterial strains optimized to generate specific bioproducts. In some cases, efficient bioproduction of a target compound can be achieved by expressing relevant metabolic pathways in different microbial species. But there are also many examples where heterologously expressed metabolic pathways perform poorly due to other physiological features of a microbe that make it a non-optimal chassis (Calero and Nikel, 2019). For this reason, it is often non-trivial to re-engineer cyanobacterial metabolism for direct synthesis of a desired compound, which may stubbornly resist efforts to improve product titer (Savakis and Hellingwerf, 2015; Nagarajan et al., 2016; Lin and Pakrasi, 2019).

A modular approach for multi-species product synthesis offers the capacity to leverage species with the most compatible physiology and desirable endogenous pathways for a given biochemical transformation, thus bypassing metabolic limitations of one biological chassis. At least in theory, each member of a synthetic microbial consortium can be conceptualized as a "module" selected specifically to perform functions well-suited with organism's abilities. In this context, cyanobacteria-heterotroph co-cultures can be rationally designed to retain the advantages of cyanobacterial metabolism (i.e., use of light/CO_2 inputs, efficient carbon fixation) and paired with other microbes that have demonstrated efficiency in transforming simple carbohydrates into a desired end product. Additionally, because the co-culture output can be changed by swapping the "heterotrophic module" (i.e., organism), some steps to optimize synthesis for one product (e.g., improving cyanobacterial sucrose production) may be transferable to achieve enhanced synthesis across many distinct cyanobacteria-heterotroph pairings. In practice, sucrose-secreting cyanobacteria have already been used as the basis for engineered microbial communities with numerous heterotrophic species and for a variety of end products (Table 2), although a number of improvements will be required to make these co-cultures feasible for scaled application.

5.2. Cyanobacterial co-culture as a flexible platform for value-added products

At the time of this writing, the most common metabolic output reported from cyanobacteria-heterotroph co-cultures are polyhydroxyalkanoates (PHAs), a class of biological polymers with comparative qualities to petroleum-based plastics. PHAs have the advantage of being both compatible in blends with commonly used petroleum-based polymers while also exhibiting superior biodegradation properties (Boey et al., 2021; Mezzina et al., 2021). Additionally, some heterotrophic microbes utilize PHAs as an intracellular storage polymer and under stress conditions can naturally hyperaccumulate PHAs in excess of 80% of their dry cell mass (Leong et al., 2014; Lee et al., 2021), making these compounds an ideal test case for the division of labor between metabolic specialists, as outlined above. Polyhydroxybutyrate (PHB) is a PHA polymer that has been produced in cyanobacterial co-culture with three different heterotrophic species: *Azotobacter vinelandii*, *Halomonas boliviensis*, and *E. coli* W (Hays et al., 2017; Smith and Francis, 2017; Weiss et al., 2017; Table 2). PHB is a natural storage polymer for both *A. vinelandii* and *H. boliviensis*, while heterologous expression of the *phaCAB* operon in *E. coli* will confer PHB synthesis capability. The most productive co-cultures reported included a heterotrophic partner species that was naturally capable of PHB synthesis. Notably, the co-cultivation of *S. elongatus* PCC 7942 *cscB* with *H. boliviensis* was extended over 6 months with no organic carbon input, demonstrating that these synthetic consortia can be stable and productive over long time periods (Weiss et al., 2017).

Pseudomonas putida is a model organism that naturally accumulates medium chain length PHAs (mcl-PHAs) granules in response to starvation, primarily under low-nitrogen and high-carbon conditions (Hoffmann and Rehm, 2004). While sucrose is not naturally consumed by *P. putida*, expression of heterologous sucrose transporters and sucrose

hydrolyzing enzymes allows it to grow on sucrose as the sole carbon source (Sabri et al., 2013; Löwe et al., 2020), a strategy that has been used to enable other microbial species without native pathways to consume cyanobacterially secreted sucrose (Sabri et al., 2013; Hobmeier et al., 2020; Zhang et al., 2020). Indeed, initial reports demonstrated that *P. putida* expressing *cscAB* was capable of growing solely on sucrose provided by *S. elongatus* PCC 7942 and accumulated PHA in co-culture, though sucrose utilization was incomplete and productivities were

TABLE 2 Sucrose-based autotroph-heterotroph co-cultures and their products.

Sucrose Strain				Heterotroph Strain				
Species[a]	Genotype[b,*]	Maximum productivity[c]	Induction	Species[d]	Genotype[e]	Product[f]	Maximum productivity[g]	Reference
*Syn*7942	*cscB*	400 mg L^{-1} d^{-1}	106 mM NaCl	*B. subtilis*	–	α-amylase	not quantified	Hays et al. (2017)[†]
				E. coli W	*phaCAB*, Δ*cscR*	PHB	0.04 mg L^{-1} d^{-1}	
*Syn*7942	*cscB*	34.2 mg L^{-1} d^{-1}	N/A; physical encapsulation	*A. vinelandii*	Δ*nifL*	PHB	8 mg L^{-1} d^{-1}	Smith and Francis (2017)[‡]
*Syn*7942	*cscB*	27.4 mg L^{-1} d^{-1}	150 mM NaCl	*A. vinelandii*	Δ*nifL*	PHB	3.8% DW d^{-1}	Smith and Francis (2016)[†]
*Syn*7942	*cscB*	0.5 mg L^{-1} d^{-1}	170 mM NaCl	*H. boliviensis*	–	PHB	28.3 mg L^{-1} d^{-1}	Weiss et al. (2017)[‡]
*Syn*7942	*cscB*	102 mg L^{-1} d^{-1}	150 mM NaCl	*P. putida* EM178	*cscRABY*, Δ*nasT*	PHA	2.3 mg L^{-1} d^{-1}	Hobmeier et al. (2020)[†]
*Syn*7942	*cscB*	250 mg L^{-1} d^{-1}	150 mM NaCl	*P. putida* EM178	*cscAB*	PHA	23.8 mg L^{-1} d^{-1}	Löwe et al. (2017)
*Syn*7942	*cscB*	108 mg L^{-1} d^{-1}**	150 mM NaCl	*P. putida* EM178	*cscRABY*, Δ*nasT*	PHA	42.1 mg L^{-1} d^{-1}	Kratzl et al. (2023)
*Syn*7942	*cscB*	45 mg L^{-1} d^{-1}	100 mM NaCl	*R. glutinis*	–	DW	24.8 mg L^{-1} DW^{-1} d^{-1}	Li et al. (2017)[†]
						TFA	1.2 mg L^{-1} d^{-1}	
*Syn*2973	*cscB*	96 mg L^{-1} d^{-1}	150 mM NaCl	*E. coli* BL21(DE3)	*cscA, cscB, cscK, mcr*	3-HP	9.8 mg L^{-1} d^{-1}	Ma et al. (2022)[†]
*Therm*PKUAC	*cscB*	18.1 mg L^{-1} d^{-1}	150 mM NaCl	*E. coli* BL21(DE3)	*efe*	ethylene	0.74 mg L^{-1} d^{-1}	Cui et al. (2022)
*Syn*7942	*cscB*	10 mg L^{-1} d^{-1}			*ispS*	isoprene	0.03 mg L^{-1} d^{-1}	
*Cup*H16	sps_{6803}, spp_{6803}, *scrY*	18.1 mg L^{-1} d^{-1}	0.3% arabinose	*E. coli* W	*vioABCDE, cscABK* Δ*cscR*	violacein	4.5 mg L^{-1} d^{-1}	Nangle et al. (2020)
					crtEBIY, cscABK Δ*cscR*	β-carotene	4.8 mg L^{-1} d^{-1}	
*Syn*2973	*cscB*	0.7 mg L^{-1} d^{-1}**	150 mM NaCl	*Y. lipolytica*	*carB, carRP*	β-carotene	325 mg L^{-1} d^{-1}	Zhao et al. (2022)[‡]
				P. putida KT2440	*sfp, bpsA*	indigoidine	1.9 g L^{-1} d^{-1}	
*Syn*7942	*cscB*	263.5 mg L^{-1} d^{-1}	150 mM NaCl	*V. natriegens*	*tyr*	melanin	1.56 mg L^{-1} d^{-1}	Li C. et al. (2022)[†]
					tal	*p*-coumaric acid	8.75 mg L^{-1} d^{-1}	
					budABC	2,3-butanediol	60 mg L^{-1} d^{-1}	
					ldh	lactate	100 mg L^{-1} d^{-1}	
Syn6803	*cscB*, sps_{6803}, Δ*ggpS*	164.3 mg L^{-1} d^{-1}	400 mM NaCl	*E. coli* W	Δ*cscR*, Inv, Parvi	ε-caprolactone	102.7 mg L^{-1} h^{-1}	Toth et al. (2022)
*Syn*7942	*cscB, sh3l*	108 mg L^{-1} d^{-1}	50 mM NaCl	*P. putida* S12	*cscA, hmfH, sh3d*	FDCA	~100% in 4 days	Lin T. Y. et al. (2020)[†,‡]
*Syn*7942	*cscB*	240 mg L^{-1} d^{-1}	100 mM NaCl	*P. putida* EM173	*cscRABY, dnt*	2,4-DNT degradation	22.7 mg L^{-1} d^{-1}	Fedeson et al. (2020)[‡]
						PHA	5.1 mg L^{-1} d^{-1}	

(Continued)

TABLE 2 (Continued)

Sucrose Strain				Heterotroph Strain				
Species[a]	Genotype[b,*]	Maximum productivity[c]	Induction	Species[d]	Genotype[e]	Product[f]	Maximum productivity[g]	Reference
Syn7942	*cscB*, sps_{7942}	200 mg L^{-1} d^{-1}	N/A	*E. coli* ATCC 8739	*ΔpflB, ΔfrdABCD, ΔmgsA, ΔnarG, ΔnapA, ΔnarZ, cscB, gtfA*	electricity	380 μW***	Zhu et al. (2022)
				S. oneidensis	*ΔnapA, glk, cscAKB*			
				G. sulfurreducens	–			

[a]*Cup*H16, *Cupriavidus necator* H16; *Syn*7942, *Synechococcus elongatus* PCC 7942; *Syn*2973, *Synechococcus elongatus* UTEX 2973; *Syn*6803, *Synechocystis* sp. PCC 6803; *Therm*PKUAC, *Thermosynechococcus elongatus* PKUAC-SCTE542.

[b]*cscB*, sucrose permease; *scrY*, sucrose porin; *sh3l*, SH3 ligand; *sps*, sucrose phosphate synthase; *spp*, sucrose phosphate phosphatase.

[c]Approximated values from axenic cultivations in conditions most similar to co-culture conditions.

[d]*A. vinelandii*, *Azotobacter vinelandii* AV3; *B. subtilis*, *Bacillus subtilis* 168; *E. coli*, *Escherichia coli*; *G. sulfurreducens*, *Geobacter sulfurrenducens* PCA; *H. boliviensis*, *Halomonas boliviensis*; *P. putida*, *Pseudomonas putida*; *R. glutinis*, *Rhodotorula glutinis*; *S. onedensis*, *Shewanella onedensis* MR-1; *Y. lipolytica, Yarrowia lipolytica* CLIB138; *V. natriegens*, *Vibrio natriegens*.

[e]*bpsA*, non-ribosomal peptide synthetase; *budABC*, 2,3-butanediol gene cluster; *carB*, phytoene dehydrogenase; *carRP*, bifunctional lycopene cyclase/phytoene synthase; *crtEBIY*, β-carotene biosynthesis cassette; *cscA*, sucrose hydrolase; *cscB*, sucrose permease; *cscK*, fructokinase; *cscR*, sucrose operon repressor; *cscY*, sucrose porin; *dnt*, dinitrotoluence degradation gene cluster; *efe*, ethylene-forming protein; *frdABCD*, operon encoding fumarate reductase; *glk*, glucokinase; *gtfA,* sucrose phosphorylase; *hmfH*, HMF/furfural oxidoreductase; Inv, *cscA* invertase gene with an N-terminal *pelB* leader sequence; *ispS*, isoprene synthase; *ldh*, D-lactate dehydrogenase; *mcr*, malonyl-CoA reductase; *mgsA*, methylglyoxal synthase; *narG*, *napA*, and *narZ,* nitrate reductases; *nasT*, nitrate response regulator; *nifL*, negative regulator of nitrogen fixation; Parvi, synthetic Baeyer–Villiger monooxygenase; *pflB,* pyruvate formate-lyase B; *phaCAB*, polyhydroxybutyrate synthesis operon; *sfp*, phosphopantetheinyl transferase; *sh3d,* SH3 domain; *tal*, tyrosine ammonia lyase; *tyr*, tyrosinase; *vioABCDE*, violacein biosynthesis cassette. Unless otherwise denoted by "Δ," genes are heterologously expressed.

[f]3-HP, 3-hydroxypropionic acid; DNT, dinitrotoluene; DW, cyanobacterial biomass dry weight; FDCA, 2,5-furandicarboxylic acid; PHA, polyhydroxyalkanoate; PHB, polyhydroxybutyrate; TFA, cyanobacterial total fatty acids.

[g]Approximated values provided or calculated from titers.

[†]Enhanced photoautotroph growth in co-culture.

[‡]Implemented spatial control of co-culture.

*Subscript in *sps* and *spp* indicates the strain that it comes from (i.e., 6803 for *Synechocystis* sp. PCC 6803, 7942 for *S. elongatus* PCC 7942).

**Values from axenic cultivation prior to the introduction of heterotroph.

***Maximum power output reported for the four-species consortium.

modest (Löwe et al., 2017; Fedeson et al., 2020). Additional expression of a sucrose porin (*cscY*) and a sucrose operon repressor (*cscR*) further improved sucrose utilization (Löwe et al., 2020), while further optimization of the nitrogen-deficiency response pathway (Hobmeier et al., 2020) and culture conditions could boost PHA titer further (Kratzl et al., 2023; Table 2).

Other co-culture products include the metabolites ethylene, isoprene, 3-hydroxypropionic acid (3-HP), and 2,3-butanediol (Table 2), which are compounds in a broader class of industrially relevant precursors widely used for chemical synthesis (e.g., diols, organic acids, gaseous alkenes; Cui et al., 2022; Li C. et al., 2022; Ma et al., 2022). In most of these reports, the heterotrophic microbe utilized were *E. coli* substrains, although the rapidly growing halophile *Vibrio natriegens* was able to produce a relatively high amount of 2,3-butanediol in co-culture (Li C. et al., 2022). Interestingly, co-cultures of *S. elongatus* PCC 7942 and *P. putida* designed to convert 5-hydroxymethylfurfural to 2,5-furandicarboxylic acid (FDCA), a common precursor molecule, exhibited higher efficiency when the two species were engineered to display complementary surface proteins (Lin T. Y. et al., 2020). The authors suggest that physical binding between the two species could improve metabolic exchange (Lin T. Y. et al., 2020), an intriguing strategy that may be valuable to develop further.

Beyond commodity products, several higher-value chemicals expand the metabolic repertoire of cyanobacteria-heterotroph co-cultures. The pigment industry makes routine use of a number of compounds that generate significant environmental hazards when chemical synthesis methods are used (Pereira and Alves, 2012). Biosynthetic pathways for pigment derivatives (e.g., indigoidine for the popular pigment, indigo) are being explored for more environmentally conscious pigment synthesis (Celedón and Díaz, 2021). Recently, co-cultures have been reported for the synthesis of indigoidine using the heterotroph *P. putida*, β-carotene with *E. coli* or the yeast *Yarrowia lipolytica*, and violacein by *E. coli* (Nangle et al., 2020; Zhao et al., 2022). Although most cyanobacteria-heterotroph co-cultures make use of the model laboratory strain *S. elongatus* PCC 7942, Zhao and colleagues used a sucrose-secreting variant of the fast-growing and high-light tolerant relative, *S. elongatus* UTEX 2973, in their co-culture experiments to produce indigoidine and β-carotene (Zhao et al., 2022). The cosmetic *p*-coumaric acid, is another higher-value compound useful for its antioxidant and antimicrobial properties (Boz, 2015; Boo, 2019). The biosynthetic pathway for *p*-coumaric acid was introduced into *V. natriegens* and co-cultures of these engineered strains with *S. elongatus* PCC 7942 allowed for photosynthetically driven *p*-coumaric acid production (Li C. et al., 2022). Other recent reports provide further evidence of the flexibility of this cyanobacterial co-cultivation system (see Table 2), including bioproduction of fatty acids (Li C. et al., 2022), ε-caprolactone (Toth et al., 2022), lactate (Li C. et al., 2022), and secreted enzymes (Hays et al., 2017). Additionally, some products can be used to feed downstream bacteria and develop more complex systems. A four-species consortium utilized lactate-consuming *Shewanella onedensis* to generate electricity and acetate, in which the latter was consumed by *Geobacter sulfurreducens* to produce CO_2 for *S. elongatus* PCC 7942 (Table 2; Zhu et al., 2022).

One final co-culture example was constructed more in the service of remediating an environmental toxin, rather than producing a specific byproduct (Fedeson et al., 2020). *S. elongatus* PCC 7942 was co-cultured with an engineered strain of *P. putida* expressing a pathway for 2,4-dinitrotoluene (2,4-DNT) degradation (Akkaya et al., 2018; Table 2). 2,4-DNT is an environmentally stable and toxic byproduct generated from the manufacture of polyurethane, pesticides, and explosives (Griest et al., 1995; Ju and Parales, 2010). In order to prepare co-cultures that were stable in the face of toxic levels of 2,4-DNT, it was necessary to encapsulate sucrose-secreting *S. elongatus* PCC 7942 within an alginate hydrogel, which increased the resilience of the cyanobacteria to the environmental stress without diminishing its capacity to perform photosynthesis and secrete sucrose for *P. putida* consumption (Fedeson et al., 2020). Notably, the strategy of immobilizing one or more microbial partner in a hydrogel was utilized in a number of the aforementioned co-culture experiments (Smith and Francis, 2017; Weiss et al., 2017; Li X. et al., 2022; Zhao et al., 2022), and encapsulated cyanobacterial strains exhibited increased resilience to environmental stressors relative to planktonic controls, while simultaneously maintaining or increasing per-cell sucrose secretion rates.

5.3. Co-culture as a platform to study microbial communities

Phototrophs and heterotrophs are often metabolically intertwined in natural contexts (Morris, 2015; Henry et al., 2016). For example, many marine *Prochlorococcus* species secrete organic carbon to neighboring heterotrophic partners that perform functions in detoxifying reactive oxygen species present in the open ocean (Morris et al., 2011; Braakman et al., 2017). It has been hypothesized that the natural export of sugars from *Prochlorococcus* and other cyanobacteria may prime them to engage with surrounding heterotrophs *via* cross-feeding, and potentially "outsource" the metabolic burden of synthesizing some nutritional requirements to other organisms (Werner et al., 2014; Henry et al., 2016; Braakman et al., 2017). Natural microbial communities and symbiotic relationships usually develop over evolutionary time scales and may exhibit numerous and complex cross-feeding patterns and other self-stabilizing interactions (Konopka et al., 2015). Yet, these important dynamics can be challenging to study due to the difficulty of disentangling specific mechanisms from the complex interaction networks (Ponomarova and Patil, 2015). The fact that many natural symbioses also have cyanobacterial partners that exchange fixed carbon for other microbial partner (s) has led some groups to explore synthetic cyanobacteria/heterotroph co-cultures as a possible "bottom-up" system to gain insight into complex microbial consortia.

Synthetic phototroph-heterotroph microbial consortia may represent a complementary system to study natural consortia in parallel, as they present a platform for interrogating microbial interactions that is relatively simple, genetically tractable, and experimentally tunable (Table 3; De Roy et al., 2014; Song et al., 2015). One intriguing phenomenon that recurs across several synthetic cyanobacteria/heterotroph co-cultures is an increase in the vigor or productivity of one or both partners relative to axenic controls. For instance, cyanobacterial growth was enhanced in mixed culture with several heterotrophic species (Hays et al., 2017; Li et al., 2017; Hobmeier et al., 2020; Ma et al., 2022), although the partner species were evolutionarily "naïve" to one another. Similarly, heterotrophic productivity in co-culture can be significantly higher than can be attributed to the cyanobacterially secreted sucrose (Hays et al., 2017; Cui et al., 2022; Ma et al., 2022). Conversely, when cyanobacteria are allowed to overpopulate a synthetic co-culture, heterotrophic partners may exhibit reduced viability (Hays et al., 2017). It is highly likely that some of these effects arise due to unprogrammed metabolic interactions and emergent behaviors of division of labor (Rafieenia et al., 2022), such as the generation of damaging reactive oxygen species (Hays et al., 2017; Ma et al., 2022).

Our current understanding of the emergent properties of mixed microbial communities is limited and cannot fully explain observed phenomena. Preliminary analyses and multi-omics approaches have been used to predict hidden interactions within synthetic consortia, providing insight on areas of cooperation and competition that could be validated and exploited to design more robust co-cultures (Carruthers, 2020; Zuñiga et al., 2020; Ma et al., 2022). Synthetic co-cultures also present a simpler set of variables in comparison to natural communities which may be more amenable to simulations, such as agent-based modeling, for predicting emergent behaviors in a population (Sakkos et al., 2022). Finally, additional layers of metabolic exchange can be designed into the synthetic co-culture system to experimentally probe and validate hypotheses of inter-species exchange. A notable example in this regard is multiple groups' use of the diazotroph, *A. vinelandii,* to fix atmospheric nitrogen and secrete ammonia, effectively creating a carbon-for-nitrogen exchange in co-culture with sucrose-secreting cyanobacteria (Smith and Francis, 2017; Carruthers, 2020). Taken together, the computational, systems, and genetic toolkits available for synthetic microbial consortia may lead to important insights on the dynamics of microbial exchange that would be difficult to probe in natural microbiomes.

6. Challenges and future perspectives

Cyanobacterial sucrose production exhibits considerable potential to facilitate sustainable bioproduction using light and CO_2 but could benefit from still further enhancements in productivity. Expanding into more elaborate metabolic engineering efforts guided by cyanobacterial genome-scale metabolic models might be one approach to identify other potential metabolic targets to increase sucrose yields. In addition, only one transporter has been used so far to facilitate sucrose secretion in cyanobacteria, CscB. While this transporter seems to work properly in many cyanobacterial strains (Table 1), *Synechocystis* sp. PCC 6803 is an exception (Du et al., 2013; Kirsch et al., 2018). CscB has a relatively low affinity for sucrose (Sahin-Tóth and Kaback, 2000), so alternative transporters with higher affinity or transport kinetics might be used to boost cyanobacterial sucrose export, expand the range of cyanobacterial species that can be engineered, or used to increase the uptake rates for co-cultured heterotrophs.

Although enzymes involved in cyanobacterial sucrose synthesis and degradation have been the subject of extensive study, there are still major gaps in our understanding of the function of these enzymes. Areas that contain a number of open questions for future study include: (i) the co-evolution of bidomainal SPS with and without SPP activity among cyanobacterial species; (ii) the role (s) and substrate (s) of SPP-like proteins in cyanobacteria; and (iii) the alternative roles of sucrose in cyanobacteria apart from its osmoprotective functions. Increasing the knowledge in all these areas will not only be useful to understand the regulation and evolution of different sucrose enzymes in cyanobacteria, but also to further engineer these enzymes to obtain high sucrose yields.

TABLE 3 Synthetic cyanobacteria-heterotroph microbial consortia used as a platform to study microbial interactions.

Sucrose strain[a]	Genotype[b]	Heterotroph strain[c]	Genotype[d]	Notes	Reference
*Syn*7942	*cscB*	*E. coli* K-12	AA knockouts	Utilizes metabolic modeling and experimental validation to predict co-cultivation outcomes and identify optimizable parameters.	Zuñiga et al. (2020)
		E. coli W	–		
		Y. lipolytica	SUC2		
		B. subtilis	–		
*Syn*7942	*cscB*, sps_{6803}	*A. vinelandii*	Δ*nifL*	Develops tripartite consortium with carbon-providing *S. elongatus* PCC 7942 and nitrogen-providing *A. vinelandii* to support a third microbe. Performed computational analyses to identify bottlenecks to improve cultivation conditions.	Carruthers (2020)
		E. coli K-12 MG1655	*cscABK*		
		C. glutamicum	–		
		B. subtilis 168	–		
*Syn*2973	*cscB*	*E. coli* BL21(DE3)	*cscABK*, *mcr*	Utilizes transcriptomic, proteomic, and metabolomic analyses to reveal differentially regulated pathways during co-cultivation to identify optimizable parameters to improve stability and 3-hydroxypropionic productivity.	Ma et al. (2022)
*Syn*7942	*cscB*	*E. coli* MG1655	*cscABK*	Spatially separates subpopulations with encapsulation to impart species stability while still allowing the transport of small molecules.	Wang et al. (2022)
*Syn*7942	*cscB*, sps_{6803}	*E. coli* W	Δ*cscR*	Utilizes individual-based modeling in spatial context to predict colony fitness.	Sakkos et al. (2022)
*Syn*7942	*cscB*, sps_{6803}	*E. coli* W	Δ*cscR*	Integrates quorum sensing modules for cross-species communication.	Kokarakis et al. (2022)

[a]*Syn*7942, *Synechococcus elongatus* PCC 7942; *Syn*2973, *Synechococcus elongatus* UTEX 2973.
[b]AA knockouts, multiple one-way amino acid auxotrophs were generated; *cscB*, sucrose permease; sps_{6803}, sucrose phosphate synthase from *Synechocystis* sp. PCC 6803.
[c]*A. vinelandii*, *Azotobacter vinelandii* AZBB163; *B. subtilis*, *Bacillus subtilis* 168; *C. glutamicum*, *Corynebacterium glutamicum* 13032; *E. coli*, *Escherichia coli*, *Y. lipolytica*, *Yarrowia lipolytica* Po1g.
[d]*cscABK*, sucrose utilization operon; *cscR*, sucrose operon repressor; *mcr*, malonyl-CoA reductase; *nifL*, negative regulator of nitrogen fixation; SUC2, cassette for internal and external invertases.

Cyanobacteria hold considerable potential as cell factories to produce sucrose, yet the development of commercially viable applications of this strategy will require a significant amount of additional research and optimization. Importantly, while yields of sucrose from cyanobacteria could theoretically exceed production from traditional plant crops at scale, significant barriers to translate results from the lab to the field are evident. For instance, deployment of outdoor cultivation would require strains that exhibit resilience to the dynamic fluctuations of temperature, light, diurnal cycles, and abiotic stresses (Jaiswal et al., 2022). Furthermore, while the bioavailability of sucrose lends itself to a high degree of flexibility in the design of co-cultures, it also makes cyanobacterial cultures highly vulnerable to invasive microbes (Hays and Ducat, 2015; Gao et al., 2022). Contamination that reduces culture output would be highly likely in any scaled system without the implementation of aggressive confinement and/or pesticidal treatments that would greatly increase the cost of production. Alternatively, efficient, automated, economical, and sustainable systems to separate secreted sucrose might be employed, as explored in a recent membrane-filtration system (Hao et al., 2022).

A much tighter integration of the signaling and metabolic exchanges between cyanobacterial and heterotroph co-culture partners might suppress contaminating species through competition and exclusion. Adaptative laboratory evolution could be a useful strategy to domesticate increasingly stable co-cultures by better integrating and adapting the partners to one another (Konstantinidis et al., 2021). Rational engineering strategies to generate more intricate coordination of

activities between species and at the population level might also contribute to this goal (Kokarakis et al., 2022). Cyanobacterial and heterotrophic partner species that have been more extensively designed to cooperate and coordinate would also be likely to exhibit higher end-product titers relative to the current productivities achievable from co-culture. The exploration of mechanisms that promote partner coordination in synthetic communities through rational and directed research efforts could provide additional insights into the underlying organizational principles in robust cyanobacterial symbioses that occupy many natural ecological niches.

7. Methods

7.1. Sequence homology

The protein sequences of the orthologues of the different enzymes involved in the sucrose biosynthesis and degradation were obtained from NCBI and Uniprot databases. These sequences were retrieved using BLAST tools in both databases; enzymes with a well-established role in these pathways in cyanobacteria were used a queries: SPS unidomainal from *Nostoc* sp. PPC 7120 (GenBank accession No. BAB76075.1; Cumino et al., 2002), SPS bidomainal from *S. elongatus* PCC 7942 (GenBank accession No. ABB56840.1; Liang et al., 2020), SPP from *Synechocystis* sp. PCC 6803 (GenBank accession No. BAA18419.1; Fieulaine et al., 2005), SuS from *Nostoc* sp. PCC 7120 (GenBank accession No. BAB76684.1; Ehira et al., 2014), AMS from *Synechococcus* sp. PCC 7002 (GenBank accession No. ACA98889.1; Perez-Cenci and Salerno, 2014) and from *Alteromonas macleodii* KCTC 2957 (Kolman and Salerno, 2016), and INV from *S. elongatus* PCC 7942 (GenBank accession No. ABB56429.1; Liang et al., 2020). For SPP-like proteins, we seeded the analysis using the sequence from *S. elongatus* PCC 7942 (GenBank accession No. ABB56598.1). Each protein in this query list was used to search for homolog sequences in the genome of 121 cyanobacterial genomes, and hits with an E-value less than or equal to 10^{-15}, an identity less than or equal to 35% and a coverage less than or equal to 80% were considered true homologs.

7.2. Multiple sequence alignments

Multiple sequence alignment analyses were performed using MEGA X (Kumar et al., 2018) and visualized with the Jalview multiple sequence alignment editor using the color scheme from ClustalX (Waterhouse et al., 2009). Logos for the conserved motifs for each analyzed enzyme were obtained using WebLogo server (Crooks et al., 2004).

7.3. Phylogenetic trees

Unrooted neighbor-joining phylogenetic trees were generated using MEGA X after the multiple sequence alignments of the sequence of SPP proteins and SPP-like proteins using ClustalX with a BLOSSUM matrix and a bootstrap trial of 1,000. The graphical representations of the trees were created using FigTree. The neighbor-joining tree of SPP-like and SPP sequences was generated using the *p*-distance substitution method including both transitions and transversions, uniform rates among sites, and pairwise deletion treatment. Support for each node was tested with 1,000 bootstrap replicates.

7.4. Protein structure analysis

The previously published crystal structures of SPS from *Thermosynechococcus vestivus* (Li et al., 2020) and SPP from *Synechocystis* sp. PCC 6803 (Fieulaine et al., 2005) were downloaded from PDB (Berman et al., 2003) with IDs 6KIH and 1U2T, respectively. All structure figures were prepared using ChimeraX (Pettersen et al., 2021).

Author contributions

MS-M, LY, and DD outlined the scope and content of the manuscript. MS-M and LY conducted the literature review and wrote the draft manuscript. MS-M conducted the phylogenetic analyses and prepared Figures 1–4 and Supplementary Table S1. LY prepared the Tables within the main manuscript, while MS-M prepared the tables in the Supplemental material. MS-M, LY, and DD reviewed, edited, and proofed the manuscript. All authors contributed to the article and approved the submitted version.

Funding

This work was primarily supported by the Department of Energy and Basic Energy Sciences Division (Grant: DE-FG02-91ER20021), and the National Science Foundation and the Division of Molecular and Cellular Bioscience (Grant: 1845463). LY was supported by a fellowship from the Plant Biotechnology for Health and Sustainability Training Program at Michigan State University (Grant: NIH T32-GM110523).

Acknowledgments

We would like to thank our laboratory postdoctoral fellows Amit Singh and Sreeahila Retnadhas, and our predoctoral fellows Emmanuel Kokarakis and Rees Rillema for helpful comments on this manuscript.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

Supplementary material

The Supplementary material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fmicb.2023.1126032/full#supplementary-material

References

Abramson, B. W., Kachel, B., Kramer, D. M., and Ducat, D. C. (2016). Increased photochemical efficiency in cyanobacteria via an engineered sucrose sink. *Plant Cell Physiol.* 57, 2451–2460. doi: 10.1093/pcp/pcw169

Abramson, B. W., Lensmire, J., Lin, Y.-T., Jennings, E., and Ducat, D. C. (2018). Redirecting carbon to bioproduction via a growth arrest switch in a sucrose-secreting cyanobacterium. *Algal Res.* 33, 248–255. doi: 10.1016/j.algal.2018.05.013

Adomako, M., Ernst, D., Simkovsky, R., Chao, Y. Y., Wang, J., Fang, M., et al. (2022). Comparative genomics of *Synechococcus elongatus* explains the phenotypic diversity of the strains. *mBio* 13:e0086222. doi: 10.1128/mbio.00862-22

Aikens, J., and Turner, R.J. (2013). *Method of producing a fermentable sugar*. Washington, DC: U.S. Patent and Trademark Office.

Akkaya, O., Pérez-Pantoja, D. R., Calles, B., Nikel, P. I., and De Lorenzo, V. (2018). The metabolic redox regime of *Pseudomonas putida* tunes its evolvability toward novel xenobiotic substrates. *mBio* 9, e01512–e01518. doi: 10.1128/mBio.01512-18

Albi, T., Ruiz, M. T., De Los Reyes, P., Valverde, F., and Romero, J. M. (2016). Characterization of the sucrose phosphate phosphatase (SPP) isoforms from *Arabidopsis thaliana* and role of the S6PPc domain in dimerization. *PLoS One* 11:e0166308. doi: 10.1371/journal.pone.0166308

Allahverdiyeva, Y., Isojarvi, J., Zhang, P., and Aro, E. M. (2015). Cyanobacterial oxygenic photosynthesis is protected by flavodiiron proteins. *Life* 5, 716–743. doi: 10.3390/life5010716

Allahverdiyeva, Y., Mustila, H., Ermakova, M., Bersanini, L., Richaud, P., Ajlani, G., et al. (2013). Flavodiiron proteins Flv1 and Flv3 enable cyanobacterial growth and photosynthesis under fluctuating light. *Proc. Natl. Acad. Sci. U. S. A.* 110, 4111–4116. doi: 10.1073/pnas.1221194110

Angermayr, S. A., Paszota, M., and Hellingwerf, K. J. (2012). Engineering a cyanobacterial cell factory for production of lactic acid. *Appl. Environ. Microbiol.* 78, 7098–7106. doi: 10.1128/AEM.01587-12

Aravind, L., Galperin, M. Y., and Koonin, E. V. (1998). The catalytic domain of the P-type ATPase has the haloacid dehalogenase fold. *Trends Biochem. Sci.* 23, 127–129. doi: 10.1016/S0968-0004(98)01189-X

Arias, D. M., Ortiz-Sanchez, E., Okoye, P. U., Rodriguez-Rangel, H., Balbuena Ortega, A., Longoria, A., et al. (2021). A review on cyanobacteria cultivation for carbohydrate-based biofuels: cultivation aspects, polysaccharides accumulation strategies, and biofuels production scenarios. *Sci. Total Environ.* 794:148636. doi: 10.1016/j.scitotenv.2021.148636

Baran, R., Lau, R., Bowen, B. P., Diamond, S., Jose, N., Garcia-Pichel, F., et al. (2017). Extensive turnover of compatible solutes in cyanobacteria revealed by deuterium oxide (D_2O) stable isotope probing. *ACS Chem. Biol.* 12, 674–681. doi: 10.1021/acschembio.6b00890

Berman, H., Henrick, K., and Nakamura, H. (2003). Announcing the worldwide protein data Bank. *Nat. Struct. Biol.* 10:980. doi: 10.1038/nsb1203-980

Blank, C. E. (2013). Phylogenetic distribution of compatible solute synthesis genes support a freshwater origin for cyanobacteria. *J. Phycol.* 49, 880–895. doi: 10.1111/jpy.12098

Blombach, B., Grunberger, A., Centler, F., Wierckx, N., and Schmid, J. (2022). Exploiting unconventional prokaryotic hosts for industrial biotechnology. *Trends Biotechnol.* 40, 385–397. doi: 10.1016/j.tibtech.2021.08.003

Boey, J. Y., Mohamad, L., Khok, Y. S., Tay, G. S., and Baidurah, S. (2021). A review of the applications and biodegradation of polyhydroxyalkanoates and poly(lactic acid) and its composites. *Polymers* 13:1544. doi: 10.3390/polym13101544

Boo, Y. C. (2019). P-Coumaric acid as an active ingredient in cosmetics: a review focusing on its antimelanogenic effects. *Antioxidants* 8:275. doi: 10.3390/antiox8080275

Boz, H. (2015). P-Coumaric acid in cereals: presence, antioxidant and antimicrobial effects. *Int. J. Food Sci. Technol.* 50, 2323–2328. doi: 10.1111/ijfs.12898

Braakman, R., Follows, M. J., and Chisholm, S. W. (2017). Metabolic evolution and the self-organization of ecosystems. *Proc. Natl. Acad. Sci. U. S. A.* 114, E3091–E3100. doi: 10.1073/pnas.1619573114

Calero, P., and Nikel, P. I. (2019). Chasing bacterial chassis for metabolic engineering: a perspective review from classical to non-traditional microorganisms. *Microb. Biotechnol.* 12, 98–124. doi: 10.1111/1751-7915.13292

Carruthers, D. (2020). *Engineering modular synthetic microbial consortia for sustainable bioproduction from CO_2*. University of Michigan.

Catone, C. M., Ripa, M., Geremia, E., and Ulgiati, S. (2021). Bio-products from algae-based biorefinery on wastewater: a review. *J. Environ. Manag.* 293:112792. doi: 10.1016/j.jenvman.2021.112792

Celedón, R. S., and Díaz, L. B. (2021). Natural pigments of bacterial origin and their possible biomedical applications. *Microorganisms* 9:739. doi: 10.3390/microorganisms9040739

Chen, B., Wan, C., Mehmood, M. A., Chang, J. S., Bai, F., and Zhao, X. (2017). Manipulating environmental stresses and stress tolerance of microalgae for enhanced production of lipids and value-added products-a review. *Bioresour. Technol.* 244, 1198–1206. doi: 10.1016/j.biortech.2017.05.170

Chua, T. K., Bujnicki, J. M., Tan, T. C., Huynh, F., Patel, B. K., and Sivaraman, J. (2008). The structure of sucrose phosphate synthase from *Halothermothrix orenii* reveals its mechanism of action and binding mode. *Plant Cell* 20, 1059–1072. doi: 10.1105/tpc.107.051193

Cid, E., Gomis, R. R., Geremia, R. A., Guinovart, J. J., and Ferrer, J. C. (2000). Identification of two essential glutamic acid residues in glycogen synthase. *J. Biol. Chem.* 275, 33614–33621. doi: 10.1074/jbc.M005358200

Collet, J. F., Stroobant, V., Pirard, M., Delpierre, G., and Van Schaftingen, E. (1998). A new class of phosphotransferases phosphorylated on an aspartate residue in an amino-terminal DXDX(T/V) motif. *J. Biol. Chem.* 273, 14107–14112. doi: 10.1074/jbc.273.23.14107

Crooks, G. E., Hon, G., Chandonia, J. M., and Brenner, S. E. (2004). WebLogo: a sequence logo generator. *Genome Res.* 14, 1188–1190. doi: 10.1101/gr.849004

Cui, Y., Rasul, F., Jiang, Y., Zhong, Y., Zhang, S., Boruta, T., et al. (2022). Construction of an artificial consortium of Escherichia coli and cyanobacteria for clean indirect production of volatile platform hydrocarbons from CO_2. *Front. Microbiol.* 13:965968. doi: 10.3389/fmicb.2022.965968

Cumino, A., Curatti, L., Giarrocco, L., and Salerno, G. L. (2002). Sucrose metabolism: *Anabaena* sucrose-phosphate synthase and sucrose-phosphate phosphatase define minimal functional domains shuffled during evolution. *FEBS Lett.* 517, 19–23. doi: 10.1016/S0014-5793(02)02516-4

Cumino, A. C., Marcozzi, C., Barreiro, R., and Salerno, G. L. (2007). Carbon cycling in anabaena sp. PCC 7120. Sucrose synthesis in the heterocysts and possible role in nitrogen fixation. *Plant Physiol.* 143, 1385–1397. doi: 10.1104/pp.106.091736

Cumino, A. C., Perez-Cenci, M., Giarrocco, L. E., and Salerno, G. L. (2010). The proteins involved in sucrose synthesis in the marine cyanobacterium *Synechococcus* sp. PCC 7002 are encoded by two genes transcribed from a gene cluster. *FEBS Lett.* 584, 4655–4660. doi: 10.1016/j.febslet.2010.10.040

Curatti, L., Flores, E., and Salerno, G. (2002). Sucrose is involved in the diazotrophic metabolism of the heterocyst-forming cyanobacterium *Anabaena* sp. *FEBS Lett.* 513, 175–178. doi: 10.1016/S0014-5793(02)02283-4

Curatti, L., Folco, E., Desplats, P., Abratti, G., Limones, V., Herrera-Estrella, L., et al. (1998). Sucrose-phosphate synthase from *Synechocystis* sp. strain PCC 6803: identification of the spsA gene and characterization of the enzyme expressed in *Escherichia coli*. *J. Bacteriol.* 180, 6776–6779. doi: 10.1128/JB.180.24.6776-6779.1998

Curatti, L., Giarrocco, L. E., Cumino, A. C., and Salerno, G. L. (2008). Sucrose synthase is involved in the conversion of sucrose to polysaccharides in filamentous nitrogen-fixing cyanobacteria. *Planta* 228, 617–625. doi: 10.1007/s00425-008-0764-7

Curatti, L., Giarrocco, L., and Salerno, G. L. (2006). Sucrose synthase and RuBisCo expression is similarly regulated by the nitrogen source in the nitrogen-fixing cyanobacterium *Anabaena* sp. *Planta* 223, 891–900. doi: 10.1007/s00425-005-0142-7

Curatti, L., Porchia, A. C., Herrera-Estrella, L., and Salerno, G. L. (2000). A prokaryotic sucrose synthase gene (susA) isolated from a filamentous nitrogen-fixing cyanobacterium encodes a protein similar to those of plants. *Planta* 211, 729–735. doi: 10.1007/s004250000343

Dan, Y., Sun, J., Zhang, S., Wu, Y., Mao, S., Luan, G., et al. (2022). Manipulating the expression of glycogen phosphorylase in *Synechococcus elongatus* PCC 7942 to mobilize glycogen storage for sucrose synthesis. *Front. Bioeng. Biotechnol.* 10:925311. doi: 10.3389/fbioe.2022.925311

Das, P., and Gundimeda, H. (2022). Is biofuel expansion in developing countries reasonable? A review of empirical evidence of food and land use impacts. *J. Clean. Prod.* 372:133501. doi: 10.1016/j.jclepro.2022.133501

De Roy, K., Marzorati, M., Van Den Abbeele, P., Van De Wiele, T., and Boon, N. (2014). Synthetic microbial ecosystems: an exciting tool to understand and apply microbial communities. *Environ. Microbiol.* 16, 1472–1481. doi: 10.1111/1462-2920.12343

Desplats, P., Folco, E., and Salerno, G. L. (2005). Sucrose may play an additional role to that of an osmolyte in *Synechocystis* sp. PCC 6803 salt-shocked cells. *Plant Physiol. Biochem.* 43, 133–138. doi: 10.1016/j.plaphy.2005.01.008

Du, W., Liang, F., Duan, Y., Tan, X., and Lu, X. (2013). Exploring the photosynthetic production capacity of sucrose by cyanobacteria. *Metab. Eng.* 19, 17–25. doi: 10.1016/j.ymben.2013.05.001

Duan, Y., Luo, Q., Liang, F., and Lu, X. (2016). Sucrose secreted by the engineered cyanobacterium and its fermentability. *J. Ocean Univ. China* 15, 890–896. doi: 10.1007/s11802-016-3007-8

Ducat, D. C., Avelar-Rivas, J. A., Way, J. C., and Silver, P. A. (2012). Rerouting carbon flux to enhance photosynthetic productivity. *Appl. Environ. Microbiol.* 78, 2660–2668. doi: 10.1128/AEM.07901-11

Ducat, D. C., Way, J. C., and Silver, P. A. (2011). Engineering cyanobacteria to generate high-value products. *Trends Biotechnol.* 29, 95–103. doi: 10.1016/j.tibtech.2010.12.003

Ehira, S., Kimura, S., Miyazaki, S., and Ohmori, M. (2014). Sucrose synthesis in the nitrogen-fixing cyanobacterium *Anabaena* sp. strain PCC 7120 is controlled by the two-component response regulator OrrA. *Appl. Environ. Microbiol.* 80, 5672–5679. doi: 10.1128/AEM.01501-14

Fedeson, D. T., Saake, P., Calero, P., Nikel, P. I., and Ducat, D. C. (2020). Biotransformation of 2,4-dinitrotoluene in a phototrophic co-culture of engineered *Synechococcus elongatus* and *Pseudomonas putida*. *Microb. Biotechnol.* 13, 997–1011. doi: 10.1111/1751-7915.13544

Fieulaine, S., Lunn, J. E., Borel, F., and Ferrer, J. L. (2005). The structure of a cyanobacterial sucrose-phosphatase reveals the sugar tongs that release free sucrose in the cell. *Plant Cell* 17, 2049–2058. doi: 10.1105/tpc.105.031229

Figueroa, C. M., Asencion Diez, M. D., Kuhn, M. L., Mcewen, S., Salerno, G. L., Iglesias, A. A., et al. (2013). The unique nucleotide specificity of the sucrose synthase from *Thermosynechococcus elongatus*. *FEBS Lett.* 587, 165–169. doi: 10.1016/j.febslet.2012.11.011

Gao, H., Manishimwe, C., Yang, L., Wang, H., Jiang, Y., Jiang, W., et al. (2022). Applications of synthetic light-driven microbial consortia for biochemicals production. *Bioresour. Technol.* 351:126954. doi: 10.1016/j.biortech.2022.126954

Gibson, R. P., Turkenburg, J. P., Charnock, S. J., Lloyd, R., and Davies, G. J. (2002). Insights into trehalose synthesis provided by the structure of the retaining glucosyltransferase OtsA. *Chem. Biol.* 9, 1337–1346. doi: 10.1016/S1074-5521(02)00292-2

Griest, W. H., Tyndall, R. L., Stewart, A. J., Caton, J. E., Vass, A. A., Ho, C. H., et al. (1995). Chemical characterization and toxicological testing of windrow composts from explosives-contaminated sediments. *Environ. Toxicol. Chem.* 14, 51–59. doi: 10.1002/etc.5620140107

Gründel, M., Scheunemann, R., Lockau, W., and Zilliges, Y. (2012). Impaired glycogen synthesis causes metabolic overflow reactions and affects stress responses in the cyanobacterium *Synechocystis* sp. PCC 6803. *Microbiology* 158, 3032–3043. doi: 10.1099/mic.0.062950-0

Hagemann, M. (2011). Molecular biology of cyanobacterial salt acclimation. *FEMS Microbiol. Rev.* 35, 87–123. doi: 10.1111/j.1574-6976.2010.00234.x

Hagemann, M. (2016). "Coping with high and variable salinity: molecular aspects of compatible solute accumulation" in *The physiology of microalgae*. eds. M. A. Borowitzka, J. Beardall and J. A. Raven (Cham: Springer International Publishing), 359–372.

Hagemann, M., and Marin, K. (1999). Salt-induced sucrose accumulation is mediated by sucrose-phosphate-synthase in cyanobacteria. *J. Plant Physiol.* 155, 424–430. doi: 10.1016/S0176-1617(99)80126-6

Hao, F., Li, X., Wang, J., Li, R., Zou, L., Wang, K., et al. (2022). Separation of bioproducts through the integration of cyanobacterial metabolism and membrane filtration: facilitating cyanobacteria's industrial application. *Membranes* 12:963. doi: 10.3390/membranes12100963

Hays, S. G., and Ducat, D. C. (2015). Engineering cyanobacteria as photosynthetic feedstock factories. *Photosynth. Res.* 123, 285–295. doi: 10.1007/s11120-014-9980-0

Hays, S. G., Yan, L. L. W., Silver, P. A., and Ducat, D. C. (2017). Synthetic photosynthetic consortia define interactions leading to robustness and photoproduction. *J. Biol. Eng.* 11:4. doi: 10.1186/s13036-017-0048-5

Henry, C. S., Bernstein, H. C., Weisenhorn, P., Taylor, R. C., Lee, J. Y., Zucker, J., et al. (2016). Microbial community metabolic modeling: a community data-driven network reconstruction. *J. Cell. Physiol.* 231, 2339–2345. doi: 10.1002/jcp.25428

Hobmeier, K., Lowe, H., Liefeldt, S., Kremling, A., and Pfluger-Grau, K. (2020). A nitrate-blind P. putida strain boosts PHA production in a synthetic mixed culture. *Front. Bioeng. Biotechnol.* 8:486. doi: 10.3389/fbioe.2020.00486

Hoffmann, N., and Rehm, B. H. (2004). Regulation of polyhydroxyalkanoate biosynthesis in *Pseudomonas putida* and *Pseudomonas aeruginosa*. *FEMS Microbiol. Lett.* 237, 1–7. doi: 10.1111/j.1574-6968.2004.tb09671.x

Huang, C. H., Shen, C. R., Li, H., Sung, L. Y., Wu, M. Y., and Hu, Y. C. (2016). CRISPR interference (CRISPRi) for gene regulation and succinate production in cyanobacterium *S. elongatus* PCC 7942. *Microb. Cell Factories* 15:196. doi: 10.1186/s12934-016-0595-3

Jaiswal, D., Sahasrabuddhe, D., and Wangikar, P. P. (2022). Cyanobacteria as cell factories: the roles of host and pathway engineering and translational research. *Curr. Opin. Biotechnol.* 73, 314–322. doi: 10.1016/j.copbio.2021.09.010

Jiang, S. Y., Chi, Y. H., Wang, J. Z., Zhou, J. X., Cheng, Y. S., Zhang, B. L., et al. (2015). Sucrose metabolism gene families and their biological functions. *Sci. Rep.* 5:17583. doi: 10.1038/s41586-020-2101-7

Ju, K. S., and Parales, R. E. (2010). Nitroaromatic compounds, from synthesis to biodegradation. *Microbiol. Mol. Biol. Rev.* 74, 250–272. doi: 10.1128/MMBR.00006-10

Juttner, F. (1983). ^{14}C-labeled metabolites in heterocysts and vegetative cells of *Anabaena cylindrica* filaments and their presumptive function as transport vehicles of organic carbon and nitrogen. *J. Bacteriol.* 155, 628–633. doi: 10.1128/jb.155.2.628-633.1983

Kageyama, H., and Waditee-Sirisattha, R. (2022). "Chapter 9 - Osmoprotectant molecules in cyanobacteria: their basic features, biosynthetic regulations, and potential applications" in *Cyanobacterial physiology*. eds. H. Kageyama and R. Waditee-Sirisattha (Massachusetts: Academic Press), 113–123.

Keshari, N., Gugger, M., Zhu, T., and Lu, X. (2019). Compatible solutes profiling and carbohydrate feedstock from diversified cyanobacteria. *Algal Res.* 43:101637. doi: 10.1016/j.algal.2019.101637

Khan, M. I., Shin, J. H., and Kim, J. D. (2018). The promising future of microalgae: current status, challenges, and optimization of a sustainable and renewable industry for biofuels, feed, and other products. *Microb. Cell Factories* 17:36. doi: 10.1186/s12934-018-0879-x

Khani-Juyabad, F., Mohammadi, P., and Zarrabi, M. (2022). Insights from cyanobacterial genomic and transcriptomic analyses into adaptation strategies in terrestrial environments. *Genomics* 114:110438. doi: 10.1016/j.ygeno.2022.110438

Kirsch, F., Klahn, S., and Hagemann, M. (2019). Salt-regulated accumulation of the compatible solutes sucrose and glucosylglycerol in cyanobacteria and its biotechnological potential. *Front. Microbiol.* 10:2139. doi: 10.3389/fmicb.2019.02139

Kirsch, F., Luo, Q., Lu, X., and Hagemann, M. (2018). Inactivation of invertase enhances sucrose production in the cyanobacterium *Synechocystis* sp. PCC 6803. *Microbiology* 164, 1220–1228. doi: 10.1099/mic.0.000708

Klähn, S., and Hagemann, M. (2011). Compatible solute biosynthesis in cyanobacteria. *Environ. Microbiol.* 13, 551–562. doi: 10.1111/j.1462-2920.2010.02366.x

Klähn, S., Steglich, C., Hess, W. R., and Hagemann, M. (2010). Glucosylglycerate: a secondary compatible solute common to marine cyanobacteria from nitrogen-poor environments. *Environ. Microbiol.* 12, 83–94. doi: 10.1111/j.1462-2920.2009.02045.x

Knoot, C. J., Ungerer, J., Wangikar, P. P., and Pakrasi, H. B. (2018). Cyanobacteria: promising biocatalysts for sustainable chemical production. *J. Biol. Chem.* 293, 5044–5052. doi: 10.1074/jbc.R117.815886

Kokarakis, E. J., Rillema, R., Ducat, D. C., and Sakkos, J. K. (2022). Developing cyanobacterial quorum sensing toolkits: toward interspecies coordination in mixed autotroph/heterotroph communities. *ACS Synth. Biol.* 12, 265–276. doi: 10.1021/acssynbio.2c00527

Kolman, M. A., Nishi, C. N., Perez-Cenci, M., and Salerno, G. L. (2015). Sucrose in cyanobacteria: from a salt-response molecule to play a key role in nitrogen fixation. *Life* 5, 102–126. doi: 10.3390/life5010102

Kolman, M. A., and Salerno, G. L. (2016). Sucrose in bloom-forming cyanobacteria: loss and gain of genes involved in its biosynthesis. *Environ. Microbiol.* 18, 439–449. doi: 10.1111/1462-2920.13071

Kolman, M. A., Torres, L. L., Martin, M. L., and Salerno, G. L. (2012). Sucrose synthase in unicellular cyanobacteria and its relationship with salt and hypoxic stress. *Planta* 235, 955–964. doi: 10.1007/s00425-011-1542-5

Konopka, A., Lindemann, S., and Fredrickson, J. (2015). Dynamics in microbial communities: unraveling mechanisms to identify principles. *ISME J.* 9, 1488–1495. doi: 10.1038/ismej.2014.251

Konstantinidis, D., Pereira, F., Geissen, E. M., Grkovska, K., Kafkia, E., Jouhten, P., et al. (2021). Adaptive laboratory evolution of microbial co-cultures for improved metabolite secretion. *Mol. Syst. Biol.* 17:e10189. doi: 10.15252/msb.202010189

Kratz, W. A., and Myers, J. (1955). Nutrition and growth of several blue-green algae. *Am. J. Bot.* 42, 282–287. doi: 10.1002/j.1537-2197.1955.tb11120.x

Kratzl, F., Kremling, A., and Pfluger-Grau, K. (2023). Streamlining of a synthetic co-culture towards an individually controllable one-pot process for polyhydroxyalkanoate production from light and CO_2. *Eng. Life Sci.* 23:e2100156. doi: 10.1002/elsc.202100156

Kumar, S., Stecher, G., Li, M., Knyaz, C., and Tamura, K. (2018). MEGA X: molecular evolutionary genetics analysis across computing platforms. *Mol. Biol. Evol.* 35, 1547–1549. doi: 10.1093/molbev/msy096

Kurniah, N. I., Sawitri, W. D., Rohman, M. S., Nugraha, Y., Hase, T., and Sugiharto, B. (2021). Mutation of UDP-glucose binding motif residues lead to increased affinity for ADP-glucose in sugarcane sucrose phosphate synthase. *Mol. Biol. Rep.* 48, 1697–1706. doi: 10.1007/s11033-021-06181-8

Lairson, L. L., Henrissat, B., Davies, G. J., and Withers, S. G. (2008). Glycosyltransferases: structures, functions, and mechanisms. *Annu. Rev. Biochem.* 77, 521–555. doi: 10.1146/annurev.biochem.76.061005.092322

Lee, J. A., Kim, H. U., Na, J.-G., Ko, Y.-S., Cho, J. S., and Lee, S. Y. (2022). Factors affecting the competitiveness of bacterial fermentation. *Trends Biotechnol.* doi: 10.1016/j.tibtech.2022.10.005

Lee, J., Park, H. J., Moon, M., Lee, J. S., and Min, K. (2021). Recent progress and challenges in microbial polyhydroxybutyrate (PHB) production from CO_2 as a sustainable feedstock: a state-of-the-art review. *Bioresour. Technol.* 339:125616. doi: 10.1016/j.biortech.2021.125616

Leong, Y. K., Show, P. L., Ooi, C. W., Ling, T. C., and Lan, J. C. (2014). Current trends in polyhydroxyalkanoates (PHAs) biosynthesis: insights from the recombinant Escherichia coli. *J. Biotechnol.* 180, 52–65. doi: 10.1016/j.jbiotec.2014.03.020

Li, X., Ding, M., Wang, M., Yang, S., Ma, X., Hu, J., et al. (2022). Proteome profiling reveals changes in energy metabolism, transport and antioxidation during drought stress in Nostoc flagelliforme. *BMC Plant Biol.* 22:162. doi: 10.1186/s12870-022-03542-8

Li, T., Li, C. T., Butler, K., Hays, S. G., Guarnieri, M. T., Oyler, G. A., et al. (2017). Mimicking lichens: incorporation of yeast strains together with sucrose-secreting cyanobacteria improves survival, growth, ROS removal, and lipid production in a stable mutualistic co-culture production platform. *Biotechnol. Biofuels* 10:55. doi: 10.1186/s13068-017-0736-x

Li, C., Wang, R., Wang, J., Liu, L., Li, H., Zheng, H., et al. (2022). A highly compatible phototrophic community for carbon-negative biosynthesis. *Angew. Chem.* 62:e202215013. doi: 10.1002/anie.202215013

Li, Y., Yao, Y., Yang, G., Tang, J., Ayala, G. J., Li, X., et al. (2020). Co-crystal structure of *Thermosynechococcus elongatus* sucrose phosphate synthase with UDP and sucrose-6-phosphate provides insight into its mechanism of action involving an oxocarbenium ion and the glycosidic bond. *Front. Microbiol.* 11:1050. doi: 10.3389/fmicb.2020.01050

Liang, C., Zhang, X., Chi, X., Guan, X., Li, Y., Qin, S., et al. (2011). Serine/threonine protein kinase SpkG is a candidate for high salt resistance in the unicellular cyanobacterium *Synechocystis* sp. PCC 6803. *PLoS One* 6:e18718. doi: 10.1371/journal.pone.0018718

Liang, Y., Zhang, M., Wang, M., Zhang, W., Qiao, C., Luo, Q., et al. (2020). Freshwater cyanobacterium *Synechococcus elongatus* PCC 7942 adapts to an environment with salt stress via ion-induced enzymatic balance of compatible solutes. *Appl. Environ. Microbiol.* 86:e02904-19. doi: 10.1128/AEM.02904-19

Lin, P. C., and Pakrasi, H. B. (2019). Engineering cyanobacteria for production of terpenoids. *Planta* 249, 145–154. doi: 10.1007/s00425-018-3047-y

Lin, T. Y., Wen, R. C., Shen, C. R., and Tsai, S. L. (2020). Biotransformation of 5-hydroxymethylfurfural to 2,5-furandicarboxylic acid by a syntrophic consortium of engineered *Synechococcus elongatus* and *Pseudomonas putida*. *Biotechnol. J.* 15:e1900357. doi: 10.1002/biot.201900357

Lin, P. C., Zhang, F., and Pakrasi, H. B. (2020). Enhanced production of sucrose in the fast-growing cyanobacterium *Synechococcus elongatus* UTEX 2973. *Sci. Rep.* 10:390. doi: 10.1038/s41598-019-57319-5

Liu, Y., and Nielsen, J. (2019). Recent trends in metabolic engineering of microbial chemical factories. *Curr. Opin. Biotechnol.* 60, 188–197. doi: 10.1016/j.copbio.2019.05.010

López-Igual, R., Flores, E., and Herrero, A. (2010). Inactivation of a heterocyst-specific invertase indicates a principal role of sucrose catabolism in heterocysts of *Anabaena* sp. *J. Bacteriol.* 192, 5526–5533. doi: 10.1128/JB.00776-10

Löwe, H., Hobmeier, K., Moos, M., Kremling, A., and Pfluger-Grau, K. (2017). Photoautotrophic production of polyhydroxyalkanoates in a synthetic mixed culture of *Synechococcus elongatus* cscB and *Pseudomonas putida* cscAB. *Biotechnol. Biofuels* 10:190. doi: 10.1186/s13068-017-0875-0

Löwe, H., Sinner, P., Kremling, A., and Pfluger-Grau, K. (2020). Engineering sucrose metabolism in *Pseudomonas putida* highlights the importance of porins. *Microb. Biotechnol.* 13, 97–106. doi: 10.1111/1751-7915.13283

Lu, Z., Dunaway-Mariano, D., and Allen, K. N. (2005). HAD superfamily phosphotransferase substrate diversification: structure and function analysis of HAD subclass IIB sugar phosphatase BT4131. *Biochemistry* 44, 8684–8696. doi: 10.1021/bi050009j

Luan, G., Zhang, S., Wang, M., and Lu, X. (2019). Progress and perspective on cyanobacterial glycogen metabolism engineering. *Biotechnol. Adv.* 37, 771–786. doi: 10.1016/j.biotechadv.2019.04.005

Lunn, J. E. (2002). Evolution of sucrose synthesis. *Plant Physiol.* 128, 1490–1500. doi: 10.1104/pp.010898

Lunn, J. E., and Ap Rees, T. (1990). Apparent equilibrium constant and mass-action ratio for sucrose-phosphate synthase in seeds of *Pisum sativum*. *Biochem. J.* 267, 739–743. doi: 10.1042/bj2670739

Lunn, J. E., Ashton, A. R., Hatch, M. D., and Heldt, H. W. (2000). Purification, molecular cloning, and sequence analysis of sucrose-6^{F}-phosphate phosphohydrolase from plants. *Proc. Natl. Acad. Sci. U. S. A.* 97, 12914–12919. doi: 10.1073/pnas.230430197

Lunn, J. E., Price, G. D., and Furbank, R. T. (1999). Cloning and expression of a prokaryotic sucrose-phosphate synthase gene from the cyanobacterium *Synechocystis* sp. PCC 6803. *Plant Mol. Biol.* 40, 297–305. doi: 10.1023/A:1006130802706

Ma, J., Guo, T., Ren, M., Chen, L., Song, X., and Zhang, W. (2022). Cross-feeding between cyanobacterium *Synechococcus* and *Escherichia coli* in an artificial autotrophic-heterotrophic coculture system revealed by integrated omics analysis. *Biotechnol. Biofuels Bioproduct.* 15:69. doi: 10.1186/s13068-022-02163-5

Ma, P., Zhang, X., Chen, L., Zhao, Q., Zhang, Q., Hua, X., et al. (2020). Comparative analysis of sucrose phosphate synthase (SPS) gene family between *Saccharum officinarum* and *Saccharum spontaneum*. *BMC Plant Biol.* 20:422. doi: 10.1186/s12870-020-02599-7

Mackay, M. A., Norton, R. S., and Borowitzka, L. J. (1984). Organic osmoregulatory solutes in cyanobacteria. *Microbiology* 130, 2177–2191. doi: 10.1099/00221287-130-9-2177

Marin, K., Huckauf, J., Fulda, S., and Hagemann, M. (2002). Salt-dependent expression of glucosylglycerol-phosphate synthase, involved in osmolyte synthesis in the cyanobacterium *Synechocystis* sp. strain PCC 6803. *J. Bacteriol.* 184, 2870–2877. doi: 10.1128/JB.184.11.2870-2877.2002

Martinez-Noël, G. M., Cumino, A. C., Kolman Mde, L., and Salerno, G. L. (2013). First evidence of sucrose biosynthesis by single cyanobacterial bimodular proteins. *FEBS Lett.* 587, 1669–1674. doi: 10.1016/j.febslet.2013.04.012

Mezzina, M. P., Manoli, M. T., Prieto, M. A., and Nikel, P. I. (2021). Engineering native and synthetic pathways in *Pseudomonas putida* for the production of tailored polyhydroxyalkanoates. *Biotechnol. J.* 16:e2000165. doi: 10.1002/biot.202000165

Morone, J., Alfeus, A. A., Vasconcelos, V., and Martins, R. (2019). Revealing the potential of cyanobacteria in cosmetics and cosmeceuticals - a new bioactive approach. *Algal Res.* 41:101541. doi: 10.1016/j.algal.2019.101541

Moronta-Barrios, F., Espinosa, J., and Contreras, A. (2013). Negative control of cell size in the cyanobacterium *Synechococcus elongatus* PCC 7942 by the essential response regulator RpaB. *FEBS Lett.* 587, 504–509. doi: 10.1016/j.febslet.2013.01.023

Morris, J. J. (2015). Black queen evolution: the role of leakiness in structuring microbial communities. *Trends Genet.* 31, 475–482. doi: 10.1016/j.tig.2015.05.004

Morris, J. J., Johnson, Z. I., Szul, M. J., Keller, M., and Zinser, E. R. (2011). Dependence of the cyanobacterium Prochlorococcus on hydrogen peroxide scavenging microbes for growth at the ocean's surface. *PLoS One* 6:e16805. doi: 10.1371/journal.pone.0016805

Nadeem, H., Rashid, M. H., Siddique, M. H., Azeem, F., Muzammil, S., Javed, M. R., et al. (2015). Microbial invertases: a review on kinetics, thermodynamics, physiochemical properties. *Process Biochem.* 50, 1202–1210. doi: 10.1016/j.procbio.2015.04.015

Nagarajan, D., Lee, D. J., Kondo, A., and Chang, J. S. (2016). Recent insights into biohydrogen production by microalgae - from biophotolysis to dark fermentation. *Bioresour. Technol.* 227, 373–387. doi: 10.1016/j.biortech.2016.12.104

Nangle, S. N., Ziesack, M., Buckley, S., Trivedi, D., Loh, D. M., Nocera, D. G., et al. (2020). Valorization of CO_2 through lithoautotrophic production of sustainable chemicals in *Cupriavidus necator*. *Metab. Eng.* 62, 207–220. doi: 10.1016/j.ymben.2020.09.002

Niederholtmeyer, H., Wolfstadter, B. T., Savage, D. F., Silver, P. A., and Way, J. C. (2010). Engineering cyanobacteria to synthesize and export hydrophilic products. *Appl. Environ. Microbiol.* 76, 3462–3466. doi: 10.1128/AEM.00202-10

Nürnberg, D. J., Mariscal, V., Bornikoel, J., Nieves-Morión, M., Krauss, N., Herrero, A., et al. (2015). Intercellular diffusion of a fluorescent sucrose analog via the septal junctions in a filamentous cyanobacterium. *mBio* 6:e02109. doi: 10.1128/mBio.02109-14

Ogawa, T., Bao, D. H., Katoh, H., Shibata, M., Pakrasi, H. B., and Bhattacharyya-Pakrasi, M. (2002). A two-component signal transduction pathway regulates manganese homeostasis in Synechocystis 6803, a photosynthetic organism. *J. Biol. Chem.* 277, 28981–28986. doi: 10.1074/jbc.M204175200

Ortiz-Reyes, E., and Anex, R. P. (2022). Economic and environmental performance of non-cellulosic fermentable carbohydrates production for biofuels and chemicals. *J. Clean. Prod.* 353:131526. doi: 10.1016/j.jclepro.2022.131526

Pade, N., and Hagemann, M. (2014). Salt acclimation of cyanobacteria and their application in biotechnology. *Life* 5, 25–49. doi: 10.3390/life5010025

Page-Sharp, M., Behm, C. A., and Smith, G. D. (1999). Involvement of the compatible solutes trehalose and sucrose in the response to salt stress of a cyanobacterial *Scytonema* species isolated from desert soils. *Biochim. Biophys. Acta* 1472, 519–528. doi: 10.1016/S0304-4165(99)00155-5

Pereira, L., and Alves, M. (2012). "Dyes - environmental impact and remediation" in *Environmental protection strategies for sustainable development*. eds. A. Malik and E. Grohmann (Dordrecht: Springer Netherlands), 111–162.

Perez-Cenci, M., and Salerno, G. L. (2014). Functional characterization of Synechococcus amylosucrase and fructokinase encoding genes discovers two novel actors on the stage of cyanobacterial sucrose metabolism. *Plant Sci.* 224, 95–102. doi: 10.1016/j.plantsci.2014.04.003

Pettersen, E. F., Goddard, T. D., Huang, C. C., Meng, E. C., Couch, G. S., Croll, T. I., et al. (2021). UCSF ChimeraX: structure visualization for researchers, educators, and developers. *Protein Sci.* 30, 70–82. doi: 10.1002/pro.3943

Ponomarova, O., and Patil, K. R. (2015). Metabolic interactions in microbial communities: untangling the Gordian knot. *Curr. Opin. Microbiol.* 27, 37–44. doi: 10.1016/j.mib.2015.06.014

Porchia, A. C., Curatti, L., and Salerno, G. L. (1999). Sucrose metabolism in cyanobacteria: sucrose synthase from *Anabaena* sp. strain PCC 7119 is remarkably different from the plant enzymes with respect to substrate affinity and amino-terminal sequence. *Planta* 210, 34–40. doi: 10.1007/s004250050651

Porchia, A. C., and Salerno, G. L. (1996). Sucrose biosynthesis in a prokaryotic organism: presence of two sucrose-phosphate synthases in *Anabaena* with remarkable differences compared with the plant enzymes. *Proc. Natl. Acad. Sci. U. S. A.* 93, 13600–13604. doi: 10.1073/pnas.93.24.13600

Potocki De Montalk, G., Remaud-Simeon, M., Willemot, R. M., Sarcabal, P., Planchot, V., and Monsan, P. (2000). Amylosucrase from *Neisseria polysaccharea*: novel catalytic properties. *FEBS Lett.* 471, 219–223. doi: 10.1016/S0014-5793(00)01406-X

Prabha, S., Vijay, A. K., Paul, R. R., and George, B. (2022). Cyanobacterial biorefinery: towards economic feasibility through the maximum valorization of biomass. *Sci. Total Environ.* 814:152795. doi: 10.1016/j.scitotenv.2021.152795

Qiao, C., Duan, Y., Zhang, M., Hagemann, M., Luo, Q., and Lu, X. (2018). Effects of reduced and enhanced glycogen pools on salt-induced sucrose production in a sucrose-secreting strain of *Synechococcus elongatus* PCC 7942. *Appl. Environ. Microbiol.* 84:e02023-17. doi: 10.1128/AEM.02023-17

Qiao, C., Zhang, M., Luo, Q., and Lu, X. (2019). Identification of two two-component signal transduction mutants with enhanced sucrose biosynthesis in *Synechococcus elongatus* PCC 7942. *J. Basic Microbiol.* 59, 465–476. doi: 10.1002/jobm.201800676

Radakovits, R., Jinkerson, R. E., Darzins, A., and Posewitz, M. C. (2010). Genetic engineering of algae for enhanced biofuel production. *Eukaryot. Cell* 9, 486–501. doi: 10.1128/EC.00364-09

Rafieenia, R., Atkinson, E., and Ledesma-Amaro, R. (2022). Division of labor for substrate utilization in natural and synthetic microbial communities. *Curr. Opin. Biotechnol.* 75:102706. doi: 10.1016/j.copbio.2022.102706

Rajeev, L., Da Rocha, U. N., Klitgord, N., Luning, E. G., Fortney, J., Axen, S. D., et al. (2013). Dynamic cyanobacterial response to hydration and dehydration in a desert biological soil crust. *ISME J.* 7, 2178–2191. doi: 10.1038/ismej.2013.83

Reed, R. H., Richardson, D. L., Warr, S. R. C., and Stewart, W. D. P. (1984). Carbohydrate accumulation and osmotic stress in cyanobacteria. *Microbiology* 130, 1–4. doi: 10.1099/00221287-130-1-1

Reed, R. H., and Stewart, W. D. P. (1985). Osmotic adjustment and organic solute accumulation in unicellular cyanobacteria from freshwater and marine habitats. *Mar. Biol.* 88, 1–9. doi: 10.1007/BF00393037

Sabri, S., Nielsen, L. K., and Vickers, C. E. (2013). Molecular control of sucrose utilization in *Escherichia coli* W, an efficient sucrose-utilizing strain. *Appl. Environ. Microbiol.* 79, 478–487. doi: 10.1128/AEM.02544-12

Sahin-Tóth, M., and Kaback, H. R. (2000). Functional conservation in the putative substrate binding site of the sucrose permease from *Escherichia coli*. *Biochemistry* 39, 6170–6175. doi: 10.1021/bi000125g

Sakkos, J. K., Santos-Merino, M., Kokarakis, E. J., Li, B., Fuentes-Cabrera, M., Zuliani, P., et al. (2022). Predicting partner fitness based on spatial structuring in a light-driven microbial community. *bioRxiv* 2022.2009.2028.510001

Salerno, G. L., and Curatti, L. (2003). Origin of sucrose metabolism in higher plants: when, how and why? *Trends Plant Sci.* 8, 63–69. doi: 10.1016/S1360-1385(02)00029-8

Salerno, G. L., Porchia, A. A. C., Vargas, W. A., and Abdian, P. L. (2004). Fructose-containing oligosaccharides: novel compatible solutes in *Anabaena* cells exposed to salt stress. *Plant Sci.* 167, 1003–1008. doi: 10.1016/j.plantsci.2004.05.029

Santos-Merino, M., Singh, A. K., and Ducat, D. C. (2019). New applications of synthetic biology tools for cyanobacterial metabolic engineering. *Front. Bioeng. Biotechnol.* 7:33. doi: 10.3389/fbioe.2019.00033

Santos-Merino, M., Singh, A. K., and Ducat, D. C. (2021a). "Sink engineering in photosynthetic microbes" in *Cyanobacteria biotechnology*. eds. N. Jens, L. Sang, S. Gregory and H. Paul (Weinheim, Germany: WILEY-VCH), 171–209.

Santos-Merino, M., Torrado, A., Davis, G. A., Rottig, A., Bibby, T. S., Kramer, D. M., et al. (2021b). Improved photosynthetic capacity and photosystem I oxidation via heterologous metabolism engineering in cyanobacteria. *Proc. Natl. Acad. Sci. U. S. A.* 118:118. doi: 10.1073/pnas.2021523118

Sanz Smachetti, M. E., Coronel, C. D., Salerno, G. L., and Curatti, L. (2020). Sucrose-to-ethanol microalgae-based platform using seawater. *Algal Res.* 45:101733. doi: 10.1016/j.algal.2019.101733

Savakis, P., and Hellingwerf, K. J. (2015). Engineering cyanobacteria for direct biofuel production from CO_2. *Curr. Opin. Biotechnol.* 33, 8–14. doi: 10.1016/j.copbio.2014.09.007

Schilling, N., and Ehrnsperger, K. (1985). Cellular differentiation of sucrose metabolism in *Anabaena variabilis*. *Z. Naturforsch. C* 40, 776–779. doi: 10.1515/znc-1985-11-1205

Scott, S. A., Davey, M. P., Dennis, J. S., Horst, I., Howe, C. J., Lea-Smith, D. J., et al. (2010). Biodiesel from algae: challenges and prospects. *Curr. Opin. Biotechnol.* 21, 277–286. doi: 10.1016/j.copbio.2010.03.005

Shinde, S., Zhang, X., Singapuri, S. P., Kalra, I., Liu, X., Morgan-Kiss, R. M., et al. (2020). Glycogen metabolism supports photosynthesis start through the oxidative pentose phosphate pathway in cyanobacteria. *Plant Physiol.* 182, 507–517. doi: 10.1104/pp.19.01184

Shoumskaya, M. A., Paithoonrangsarid, K., Kanesaki, Y., Los, D. A., Zinchenko, V. V., Tanticharoen, M., et al. (2005). Identical Hik-Rre systems are involved in perception and transduction of salt signals and hyperosmotic signals but regulate the expression of individual genes to different extents in *Synechocystis*. *J. Biol. Chem.* 280, 21531–21538. doi: 10.1074/jbc.M412174200

Singh, A. K., Santos-Merino, M., Sakkos, J. K., Walker, B. J., and Ducat, D. C. (2022). Rubisco regulation in response to altered carbon status in the cyanobacterium *Synechococcus elongatus* PCC 7942. *Plant Physiol.* 189, 874–888. doi: 10.1093/plphys/kiac065

Skov, L. K., Mirza, O., Henriksen, A., De Montalk, G. P., Remaud-Simeon, M., Sarcabal, P., et al. (2001). Amylosucrase, a glucan-synthesizing enzyme from the alpha-amylase family. *J. Biol. Chem.* 276, 25273–25278. doi: 10.1074/jbc.M010998200

Smith, M. J., and Francis, M. B. (2016). A designed a. vinelandii-S. elongatus coculture for chemical photoproduction from air, water, phosphate, and trace metals. *ACS Synth. Biol.* 5, 955–961. doi: 10.1021/acssynbio.6b00107

Smith, M. J., and Francis, M. B. (2017). Improving metabolite production in microbial co-cultures using a spatially constrained hydrogel. *Biotechnol. Bioeng.* 114, 1195–1200. doi: 10.1002/bit.26235

Song, H. S., Renslow, R. S., Fredrickson, J. K., and Lindemann, S. R. (2015). Integrating ecological and engineering concepts of resilience in microbial communities. *Front. Microbiol.* 6:1298. doi: 10.3389/fmicb.2015.01298

Song, K., Tan, X., Liang, Y., and Lu, X. (2016). The potential of *Synechococcus elongatus* UTEX 2973 for sugar feedstock production. *Appl. Microbiol. Biotechnol.* 100, 7865–7875. doi: 10.1007/s00253-016-7510-z

Sturm, A. (1999). Invertases. Primary structures, functions, and roles in plant development and sucrose partitioning. *Plant Physiol.* 121, 1–8. doi: 10.1104/pp.121.1.1

Su, Y., Song, K., Zhang, P., Su, Y., Cheng, J., and Chen, X. (2017). Progress of microalgae biofuels' commercialization. *Renew. Sust. Energ. Rev.* 74, 402–411. doi: 10.1016/j.rser.2016.12.078

Sun, Z., Dzakpasu, M., Zhao, L., Wang, Z., Zhang, D., Qu, M., et al. (2022). Enhancement of partial denitrification-anammox pathways in constructed wetlands by plant-based external carbon sources. *J. Clean. Prod.* 370:133581. doi: 10.1016/j.jclepro.2022.133581

Suzuki, E., Ohkawa, H., Moriya, K., Matsubara, T., Nagaike, Y., Iwasaki, I., et al. (2010). Carbohydrate metabolism in mutants of the cyanobacterium *Synechococcus elongatus* PCC 7942 defective in glycogen synthesis. *Appl. Environ. Microbiol.* 76, 3153–3159. doi: 10.1128/AEM.00397-08

Tanabe, Y., Hodoki, Y., Sano, T., Tada, K., and Watanabe, M. M. (2018). Adaptation of the freshwater bloom-forming cyanobacterium Microcystis aeruginosa to brackish water is driven by recent horizontal transfer of sucrose genes. *Front. Microbiol.* 9:1150. doi: 10.3389/fmicb.2018.01150

Tanabe, Y., Yamaguchi, H., Sano, T., and Kawachi, M. (2019). A novel salt-tolerant genotype illuminates the sucrose gene evolution in freshwater bloom-forming cyanobacterium Microcystis aeruginosa. *FEMS Microbiol. Lett.* 366:fnz190. doi: 10.1093/femsle/fnz190

Tauzin, A. S., and Giardina, T. (2014). Sucrose and invertases, a part of the plant defense response to the biotic stresses. *Front. Plant Sci.* 5:293. doi: 10.3389/fpls.2014.00293

Thiel, K., Patrikainen, P., Nagy, C., Fitzpatrick, D., Pope, N., Aro, E. M., et al. (2019). Redirecting photosynthetic electron flux in the cyanobacterium *Synechocystis* sp. PCC 6803 by the deletion of flavodiiron protein Flv3. *Microb. Cell Factories* 18:189. doi: 10.1186/s12934-019-1238-2

Toth, G. S., Siitonen, V., Nikkanen, L., Sovic, L., Kallio, P., Kourist, R., et al. (2022). Photosynthetically produced sucrose by immobilized *Synechocystis* sp. PCC 6803 drives biotransformation in *E. coli*. *Biotechnol. Biofuels Bioproduct.* 15:146. doi: 10.1186/s13068-022-02248-1

Tsuge, Y., Kawaguchi, H., Sasaki, K., and Kondo, A. (2016). Engineering cell factories for producing building block chemicals for bio-polymer synthesis. *Microb. Cell Factories* 15:19. doi: 10.1186/s12934-016-0411-0

Vadyvaloo, V., Smirnova, I. N., Kasho, V. N., and Kaback, H. R. (2006). Conservation of residues involved in sugar/H(+) symport by the sucrose permease of *Escherichia coli* relative to lactose permease. *J. Mol. Biol.* 358, 1051–1059. doi: 10.1016/j.jmb.2006.02.050

Vargas, W., Cumino, A., and Salerno, G. L. (2003). Cyanobacterial alkaline/neutral invertases. Origin of sucrose hydrolysis in the plant cytosol? *Planta* 216, 951–960. doi: 10.1007/s00425-002-0943-x

Vargas, W. A., Nishi, C. N., Giarrocco, L. E., and Salerno, G. L. (2011). Differential roles of alkaline/neutral invertases in Nostoc sp. PCC 7120: Inv-B isoform is essential for diazotrophic growth. *Planta* 233, 153–162. doi: 10.1007/s00425-010-1288-5

Vargas, W. A., and Salerno, G. L. (2010). The Cinderella story of sucrose hydrolysis: alkaline/neutral invertases, from cyanobacteria to unforeseen roles in plant cytosol and organelles. *Plant Sci.* 178, 1–8. doi: 10.1016/j.plantsci.2009.09.015

Wan, H., Wu, L., Yang, Y., Zhou, G., and Ruan, Y. L. (2018). Evolution of sucrose metabolism: the dichotomy of invertases and beyond. *Trends Plant Sci.* 23, 163–177. doi: 10.1016/j.tplants.2017.11.001

Wang, W., Kim, R., Jancarik, J., Yokota, H., and Kim, S. H. (2001). Crystal structure of phosphoserine phosphatase from *Methanococcus jannaschii*, a hyperthermophile, at 1.8 A resolution. *Structure* 9, 65–71. doi: 10.1016/S0969-2126(00)00558-X

Wang, L., Lei, X., Yang, J., Wang, S., Liu, Y., and Liang, W. (2018). Comparative transcriptome analysis reveals that photosynthesis contributes to drought tolerance of *Nostoc flagelliforme* (Nostocales, Cyanobacteria). *Phycologia* 57, 113–120. doi: 10.2216/17-18.1

Wang, L., Zhang, X., Tang, C., Li, P., Zhu, R., Sun, J., et al. (2022). Engineering consortia by polymeric microbial swarmbots. *Nat. Commun.* 13:3879. doi: 10.1038/s41467-022-31467-1

Waterhouse, A. M., Procter, J. B., Martin, D. M., Clamp, M., and Barton, G. J. (2009). Jalview version 2--a multiple sequence alignment editor and analysis workbench. *Bioinformatics* 25, 1189–1191. doi: 10.1093/bioinformatics/btp033

Weiss, T. L., Young, E. J., and Ducat, D. C. (2017). A synthetic, light-driven consortium of cyanobacteria and heterotrophic bacteria enables stable polyhydroxybutyrate production. *Metab. Eng.* 44, 236–245. doi: 10.1016/j.ymben.2017.10.009

Wendisch, V. F., Jorge, J. M. P., Perez-Garcia, F., and Sgobba, E. (2016). Updates on industrial production of amino acids using *Corynebacterium glutamicum*. *World J. Microbiol. Biotechnol.* 32:105. doi: 10.1007/s11274-016-2060-1

Werner, G. D., Strassmann, J. E., Ivens, A. B., Engelmoer, D. J., Verbruggen, E., Queller, D. C., et al. (2014). Evolution of microbial markets. *Proc. Natl. Acad. Sci. U. S. A.* 111, 1237–1244. doi: 10.1073/pnas.1315980111

Whitton, B. A., and Potts, M. (2002). "Introduction to the cyanobacteria" in *The ecology of cyanobacteria: Their diversity in time and space*. eds. B. A. Whitton and M. Potts (Dordrecht: Springer Netherlands), 1–11.

Winter, H., and Huber, S. C. (2000). Regulation of sucrose metabolism in higher plants: localization and regulation of activity of key enzymes. *Crit. Rev. Biochem. Mol. Biol.* 35, 253–289. doi: 10.1080/10409230008984165

Wu, S., Snajdrova, R., Moore, J. C., Baldenius, K., and Bornscheuer, U. T. (2021). Biocatalysis: enzymatic synthesis for industrial applications. *Angew. Chem.* 60, 88–119. doi: 10.1002/anie.202006648

Xie, J., Cai, K., Hu, H. X., Jiang, Y. L., Yang, F., Hu, P. F., et al. (2016). Structural analysis of the catalytic mechanism and substrate specificity of *Anabaena* alkaline invertase InvA reveals a novel glucosidase. *J. Biol. Chem.* 291, 25667–25677. doi: 10.1074/jbc.M116.759290

Xie, J., Hu, H. X., Cai, K., Xia, L. Y., Yang, F., Jiang, Y. L., et al. (2018). Structural and enzymatic analyses of *Anabaena* heterocyst-specific alkaline invertase InvB. *FEBS Lett.* 592, 1589–1601. doi: 10.1002/1873-3468.13041

Xu, Y., Guerra, L. T., Li, Z., Ludwig, M., Dismukes, G. C., and Bryant, D. A. (2013). Altered carbohydrate metabolism in glycogen synthase mutants of *Synechococcus* sp. strain PCC 7002: cell factories for soluble sugars. *Metab. Eng.* 16, 56–67. doi: 10.1016/j.ymben.2012.12.002

Yamaguchi, K., Suzuki, I., Yamamoto, H., Lyukevich, A., Bodrova, I., Los, D. A., et al. (2002). A two-component Mn^{2+}-sensing system negatively regulates expression of the mntCAB operon in *Synechocystis*. *Plant Cell* 14, 2901–2913. doi: 10.1105/tpc.006262

Yu, J., Liberton, M., Cliften, P. F., Head, R. D., Jacobs, J. M., Smith, R. D., et al. (2015). *Synechococcus elongatus* UTEX 2973, a fast growing cyanobacterial chassis for biosynthesis using light and CO_2. *Sci. Rep.* 5:8132. doi: 10.1038/srep08132

Zhang, L., Chen, L., Diao, J., Song, X., Shi, M., and Zhang, W. (2020). Construction and analysis of an artificial consortium based on the fast-growing cyanobacterium

Synechococcus elongatus UTEX 2973 to produce the platform chemical 3-hydroxypropionic acid from CO_2. *Biotechnol. Biofuels* 13:82. doi: 10.1186/s13068-020-01720-0

Zhang, S., Sun, H., Sun, J., Luo, Q., Luan, G., and Lu, X. (2021). "Engineering cyanobacteria cell factories for photosynthetic production of sucrose" in *Ecophysiology and biochemistry of cyanobacteria*. ed. R. P. Rastogi (Singapore: Springer Nature Singapore), 373–399.

Zhao, R., Sengupta, A., Tan, A. X., Whelan, R., Pinkerton, T., Menasalvas, J., et al. (2022). Photobiological production of high-value pigments via compartmentalized co-cultures using Ca-alginate hydrogels. *Sci. Rep.* 12:22163. doi: 10.1038/s41598-022-26437-y

Zhong, C. (2020). Industrial-scale production and applications of bacterial cellulose. *Front. Bioeng. Biotechnol.* 8:605374. doi: 10.3389/fbioe.2020.605374

Zhu, H., Xu, L., Luan, G., Zhan, T., Kang, Z., Li, C., et al. (2022). A miniaturized bionic ocean-battery mimicking the structure of marine microbial ecosystems. *Nat. Commun.* 13:5608. doi: 10.1038/s41467-022-33358-x

Zorina, A., Sinetova, M. A., Kupriyanova, E. V., Mironov, K. S., Molkova, I., Nazarenko, L. V., et al. (2016). Synechocystis mutants defective in manganese uptake regulatory system, ManSR, are hypersensitive to strong light. *Photosynth. Res.* 130, 11–17. doi: 10.1007/s11120-015-0214-x

Zuñiga, C., Li, T., Guarnieri, M. T., Jenkins, J. P., Li, C. T., Bingol, K., et al. (2020). Synthetic microbial communities of heterotrophs and phototrophs facilitate sustainable growth. *Nat. Commun.* 11:3803. doi: 10.1038/s41467-020-17612-8

frontiers | Frontiers in Microbiology

TYPE Original Research
PUBLISHED 16 February 2023
DOI 10.3389/fmicb.2023.1112307

OPEN ACCESS

EDITED BY
Jin Liu,
Peking University,
China

REVIEWED BY
Jörg Soppa,
Goethe University Frankfurt,
Germany
David L. Bernick,
University of California, Santa Cruz,
United States

*CORRESPONDENCE
Wolfgang R. Hess
✉ wolfgang.hess@biologie.uni-freiburg.de

†PRESENT ADDRESS
Ute A. Hoffmann,
School of Engineering Sciences in Chemistry, Biotechnology and Health, Science for Life Laboratory, KTH - Royal Institute of Technology, Stockholm, Sweden

‡These authors share first authorship

SPECIALTY SECTION
This article was submitted to Microbiotechnology, a section of the journal Frontiers in Microbiology

RECEIVED 30 November 2022
ACCEPTED 23 January 2023
PUBLISHED 16 February 2023

CITATION
Kaltenbrunner A, Reimann V, Hoffmann UA, Aoyagi T, Sakata M, Nimura-Matsune K, Watanabe S, Steglich C, Wilde A and Hess WR (2023) Regulation of pSYSA defense plasmid copy number in *Synechocystis* through RNase E and a highly transcribed asRNA. *Front. Microbiol.* 14:1112307. doi: 10.3389/fmicb.2023.1112307

Regulation of pSYSA defense plasmid copy number in *Synechocystis* through RNase E and a highly transcribed asRNA

Alena Kaltenbrunner[1‡], Viktoria Reimann[1‡], Ute A. Hoffmann[2†], Tomohiro Aoyagi[3], Minori Sakata[3], Kaori Nimura-Matsune[3], Satoru Watanabe[3], Claudia Steglich[1], Annegret Wilde[2] and Wolfgang R. Hess[1*]

[1]Genetics and Experimental Bioinformatics Group, Faculty of Biology, University of Freiburg, Freiburg, Germany, [2]Molecular Genetics of Prokaryotes, Institute of Biology III, University of Freiburg, Freiburg, Germany, [3]Department of Bioscience, Tokyo University of Agriculture, Tokyo, Japan

Synthetic biology approaches toward the development of cyanobacterial producer strains require the availability of appropriate sets of plasmid vectors. A factor for the industrial usefulness of such strains is their robustness against pathogens, such as bacteriophages infecting cyanobacteria. Therefore, it is of great interest to understand the native plasmid replication systems and the CRISPR-Cas based defense mechanisms already present in cyanobacteria. In the model cyanobacterium *Synechocystis* sp. PCC 6803, four large and three smaller plasmids exist. The ~100kb plasmid pSYSA is specialized in defense functions by encoding all three CRISPR-Cas systems and several toxin-antitoxin systems. The expression of genes located on pSYSA depends on the plasmid copy number in the cell. The pSYSA copy number is positively correlated with the expression level of the endoribonuclease E. As molecular basis for this correlation we identified the RNase E-mediated cleavage within the pSYSA-encoded *ssr7036* transcript. Together with a cis-encoded abundant antisense RNA (asRNA1), this mechanism resembles the control of ColE1-type plasmid replication by two overlapping RNAs, RNA I and II. In the ColE1 mechanism, two non-coding RNAs interact, supported by the small protein Rop, which is encoded separately. In contrast, in pSYSA the similar-sized protein Ssr7036 is encoded within one of the interacting RNAs and it is this mRNA that likely primes pSYSA replication. Essential for plasmid replication is furthermore the downstream encoded protein Slr7037 featuring primase and helicase domains. Deletion of *slr7037* led to the integration of pSYSA into the chromosome or the other large plasmid pSYSX. Moreover, the presence of *slr7037* was required for successful replication of a pSYSA-derived vector in another model cyanobacterium, *Synechococcus elongatus* PCC 7942. Therefore, we annotated the protein encoded by *slr7037* as Cyanobacterial Rep protein A1 (CyRepA1). Our findings open new perspectives on the development of shuttle vectors for genetic engineering of cyanobacteria and of modulating the activity of the entire CRISPR-Cas apparatus in *Synechocystis* sp. PCC 6803.

KEYWORDS

antisense RNA, cyanobacteria, expression vectors, plasmid replication, CRISPR, RNase E, replication protein

Introduction

The photosynthetic *Synechocystis* sp. PCC 6803 (from here: *Synechocystis* 6803) is a widely-used unicellular model cyanobacterium. The complete sequence of its chromosome was determined as early as 1996 (Kaneko et al., 1996) making it the first phototrophic and the third organism overall for which a major part of its genome sequence became available. In addition to its circular chromosome, the four large plasmids pSYSA, pSYSG, pSYSM, pSYSX (Kaneko et al., 2003) and three smaller plasmids pCC5.2, pCA2.4 and pCB2.4 (5.2 kb, 2.4 kb, and 2.3 kb) exist in *Synechocystis* 6803, which were all eventually sequenced as well (Yang and McFadden, 1993, 1994; Xu and McFadden, 1997). Although the combined coding capacity of these seven plasmids makes up almost 10% of all protein-coding genes in *Synechocystis* 6803, their biological functions have only partially been addressed. Recently, a large gene cluster for the biosynthesis of an extracellular sulfated polysaccharide, called synechan, was identified on the ~120 kb plasmid pSYSM (Maeda et al., 2021).

Of particular interest is the plasmid pSYSA. This plasmid was previously characterized as specialized for defense functions by encoding all three CRISPR-Cas systems in *Synechocystis* 6803 (Scholz et al., 2013; Reimann et al., 2017; Behler et al., 2018) and at least seven distinct type II toxin-antitoxin systems (Kopfmann and Hess, 2013; Kopfmann et al., 2016). Resequencing of the *Synechocystis* 6803 lab strain PCC-M in 2012 revealed two deletion events in pSYSA (Trautmann et al., 2012) compared to the original sequence determined 10 years before (Kaneko et al., 2003). Both of these deletions were mapped to the CRISPR-Cas system, removing 2,399 bp and 159 bp, respectively (Trautmann et al., 2012).

While the small endogenous plasmids in *Synechocystis* 6803 replicate by the rolling circle mechanism (Yang and McFadden, 1993, 1994; Xu and McFadden, 1997), the replication type of the large plasmids and also of plasmids in other cyanobacteria have largely remained unknown so far.

Previous analyses after the transient inactivation of RNase E by temperature shift (TIER-seq) revealed that transcripts derived from the four major plasmids accumulated differentially between cells expressing an unmodified form of this riboendonuclease, *rne*(WT) and a temperature-sensitive form, *rne*(Ts) (Hoffmann et al., 2021). Moreover, the joint overexpression of RNase E and RNase HII led up to 3.8- and 2.4-fold increased copy numbers of plasmids pSYSA and pSYSM, respectively (Hoffmann et al., 2023). These results were interpreted as pointing toward a possible role of RNase E in copy number regulation of these replicons. Indeed, such functions were described for certain endoribonucleases. RNase E is involved in the control of ColE1-type plasmid replication in *Escherichia coli* (Lin-Chao and Cohen, 1991) and RNases J1 and J2 in the RepA_N-family plasmid pSA564 replication in *Staphylococcus aureus* (Guimarães et al., 2021).

Here we show that two abundant and overlapping pSYSA transcripts, the *ssr7036* mRNA and the asRNA1, are substrates for RNase E *in vivo* and *in vitro*. Manipulation of their levels and processing patterns leads to an altered plasmid copy number. Transfer of this locus to an *E. coli* pUC vector yielded a plasmid capable of self-replication in *Synechocystis* 6803 and, if the Slr7037 replication protein was present, also in the unrelated *Synechococcus elongatus* PCC 7942 (from here *Synechococcus* 7942). In addition, deletion of *slr7037* led to the integration of pSYSA into the chromosome or the other large plasmid pSYSX, suggesting that Slr7037 is crucial for pSYSA replication. This system resembles the copy control mechanism of ColE1-type plasmids, but differs upon closer inspection in several details.

Materials and methods

Culture media, strains, growth conditions and manipulation of cyanobacteria

If not stated otherwise, the *Synechocystis* 6803 strains PCC-M was used as wild type as re-sequenced in 2012 (Trautmann et al., 2012). Cultures were grown in BG-11 (Rippka et al., 1979) substituted with 0.3% (w/v) sodium thiosulfate and 10 mM N-[Tris-(hydroxy-methyl)-methyl]-2-aminoethanesulfonic acid (TES) buffer (pH 8.0). If not stated otherwise, liquid cultures were grown in Erlenmeyer flasks at 30°C under continuous white light illumination (30 µmol photons m^{-2} s^{-1} or as stated) and with constant shaking at 135 rpm. Plate cultures were grown on 0.75% bacto-agar BG-11 plates with added antibiotics as needed (kanamycin, 50 µg/ml, chloramphenicol, 10 µg/ml, gentamicin, 2 µg/ml).

Synechocystis 6803 was transformed with the plasmids VIII.22, VIII.23, VIII.44 and VIII.45 (Table 1). For one transformation, aliquots corresponding to 20 OD_{750} units were taken from exponentially growing cultures at an OD_{750} between 0.6 and 0.9. Cells were collected by centrifugation at 3,237 g for 6 min in a swing-out rotor at room temperature and resuspended in a droplet of remaining BG-11. Five µg of plasmid DNA were added. Thereafter, cells were incubated for 1.5 h in the light and then plated on 0.9% bacto-agar BG-11 plates. After 24 h, 400 µl of BG-11 supplemented with 50 µg/ml kanamycin (calculated according to the total plate volume) were pipetted below the agar. After 14 days at 30°C with a light intensity of 50 µmol photons m^{-2} s^{-1}, first colonies appeared. These were transferred onto fresh 0.75% bacto-agar BG-11 plates containing 40 µg/ml kanamycin.

The pVZ322 (Zinchenko et al., 1999) based plasmids VI.16, VIII.17, VIII.18 and VIII.58 were transferred into the *Synechocystis* cells by electroporation. Twenty-five mL of wild-type culture with an OD_{750} of 0.8–1.0 were harvested at 3,237 g for 10 min at room temperature in a swing-out rotor. The cell pellet was resuspended in 2 ml ice-cold HEPES (1 mM, pH 7.5) and the cells were collected as before. This wash step was repeated twice. Thereafter, the pellet was resuspended in 100 µl ice-cold HEPES per approach and 1 µg of plasmid DNA was added. The electrocompetent cells were transferred into ice-cold electroporation cuvettes. The electroporation was performed at 2,500 V for 4 ms in an electropulser (MicroPulser, Bio-Rad). Electroporated cells were resuspended first in 1 ml BG-11 medium, then added to 50 ml BG-11 in an Erlenmeyer flask and incubated for 24 h at 30°C and 50 µmol photons m^{-2} s^{-1}. Thereafter, cells were harvested at 3,237 g for 10 min, resuspended in a drop of the supernatant and spread onto 0.75% bacto-agar BG-11 plates supplemented with 2 µg/ml gentamycin. Colonies began to appear after 8 days.

For transformation of *Synechococcus* 7942, a 2 ml culture at an OD_{750} of 1.0 to 1.7 was harvested by centrifugation at 11,000 *g* for 2 min at room temperature. The supernatant was removed and the pellet resuspended in 1 ml fresh BG-11. The cells were centrifuged again at 11,000 *g* for 1 min at room temperature, the supernatant removed and the cells resuspended in 100 µl BG-11 by gently pipetting up and down before 5 µg of plasmid DNA were added. This mixture was incubated at 34°C for 4 h in the dark, then plated on 0.9% bacto-agar BG-11 plates. The following addition of kanamycin and the selection conditions were

TABLE 1 Strains, plasmid vectors and used constructs.

Strain	Description	Reference
E. coli DH5α	F⁻ φ80*lac*ZΔM15 Δ(*lac*ZYA-*arg*F)U169 *rec*A1 *end*A1 *hsd*R17(r_K^-, m_K^+) *pho*A *sup*E44 λ⁻*thi*-1 *gyr*A96 *rel*A1	Thermo Fisher Scientific
E. coli Top10F´	Identical to TOP10 cells, with the addition of an F´ episome	Thermo Fisher Scientific
E. coli M15 [pREP4] pQE70-*rne*	Expression of recombinant *Synechocystis* 6803 RNase E	Behler et al. (2018)
Synechocystis 6803 pVZΔKm^R	*slr1128*::*kanR 3xFLAG-rne*$^+$ pVZ321Δ(9220–1,203)	Hoffmann et al. (2021)
Synechocystis 6803 *3xFLAG-rne*	*slr1128*::*kanR 3xFLAG-rne*$^+$	Hoffmann et al. (2021)
Synechocystis 6803 *rne*(WT)	Δ(*rne-rnhB*) *rnep*::*kanR* pVZ321Δ(9,220-1,203) (9,219)::(*rnep-3xFLAG-rne*$^+$-*rnhB-rnhBt*)	Hoffmann et al. (2021)
Synechocystis 6803 *rne*(5p)	Δ(*rne-rnhB*) *rnep*::*kanR* pVZ321Δ(9,220-1,203) (9,219)::(*rnep-3xFLAG-rne*(5p)-*rnhB-rnhBt*)	Hoffmann et al. (2023)
Synechocystis 6803 Δ*slr7037*-1	Gene *slr7037* deleted by replacement with a chloramphenicol resistance cassette (pSYSA-chromosome integration *via slr7104-slr7105/ sll0699-sll0700*)	This work
Synechocystis 6803 Δ*slr7037*-2	Gene *slr7037* deleted by replacement with a chloramphenicol resistance cassette (pSYSA-chromosome integration *via slr7104-slr7105/ sll1256-sll1257*)	This work
Synechocystis 6803 Δ*slr7037*-3	Gene *slr7037* deleted by replacement with a chloramphenicol resistance cassette (pSYSA-pSYSX integration *via slr7005/ sll6109*)	This work
Plasmid	**Description**	**Reference**
pVZ322	IncQ KmR GmR Mob+; replicative vector for cyanobacteria	Zinchenko et al. (1999)
VI.16	Renamed here from pVZ322 derivative (pVZ322s) in which kanamycin resistance marker was deleted	Behler et al. (2018)
VIII.17	pVZ322s::P_{rha}-7036-3xFLAG	This work
VIII.18	pVZ322s::P_{rha}-GFP-3xFLAG	This work
VIII.22	pUC19s with inserted *ssr7036* and kanamycin resistance genes	This work
VIII.23	pUC19s backbone and the entire *ssr7036-slr7037* region including the pSYSA origin of replication	This work
VIII.44	pUC19s::nP7036-7037_P_{petE}sfGFP_KmR (VIII.23 plus sfGFP under control of P_{petE})	This work
VIII.45	pUC19s::nP7036-7037STOP_KmR	This work
VII.48	pUC19::T1-P_{rha}-RBS-eYFP-3xFLAG-T2-KmR-RhaS-T3 amplified from pCK355	Kelly et al. (2019) and Luisa Hemm, unpublished
VIII.58	pVZ322s::P_{rha}-7036STOP-3xFLAG (VIII.17 with 2nd codon modified to STOP codon in *ssr7036*)	This work
pUC19	GenBank entry L09137.2	Yanisch-Perron et al. (1985)
pUC19s	pUC19 backbone of 1,965 bp lacking nt 26 to 746 from GenBank entry L09137.2	This work

identical to the procedure described for the transformation of *Synechocystis* 6803. After 8 days first transformed colonies appeared.

Construction of mutant strains

All vectors and constructs used in this study are listed in Table 1. Construction of strains *rne*(WT), *rne*(Ts) and *rne*(5p) were previously described (Hoffmann et al., 2021, 2023). For the construction of plasmids VIII.22 and VIII.23, the primers P1 and P2 were used to inversely amplify the pUC19 backbone (for primer details see Table 2). The gene *ssr7036* including its 5′ and 3′ untranslated regions (UTRs) was amplified using the primers P3 and P4 and *Synechocystis* 6803 genomic DNA as template. Genes *ssr7036* and *slr7037* including the intergenic region and their 5′ and 3′ UTRs were amplified with primers P3 and P5. The plasmids were constructed using AQUA cloning (Beyer et al., 2015) and chemocompetent *E. coli* DH5α cells (Table 1). The plasmid VIII.44 was constructed using primers P2 and P6 to amplify the VIII.23 plasmid as backbone, primers P7 and P8 to amplify the *petE* promoter (P_{petE}) and primers P9 and P10 to amplify the *sfGFP* gene from plasmid VII.65. The plasmid VIII.45 is identical to VIII.23 except for a point mutation in *slr7037* leading to an early stop codon. This point mutation was inserted through site-directed mutagenesis, using the primers P11 and P12 with the VIII.23 plasmid as a template.

Plasmid VIII.17 is based on plasmid VI.16 with the following insert: T1-P_{Rha}-RBS-eYFP-3xFLAG-T2-*KmR*-RhaS-T3 that was amplified from VII.48. T1-T3 designate the *ilvBN*, ECK120034435[1] and bacteriophage λ OOP rho-independent transcriptional terminators (Dühring et al., 2006; Kelly et al., 2019). The plasmid VII.48 was inverse PCR amplified excluding eYFP, using primers P13 and P14. The gene *ssr7036* was PCR amplified from genomic DNA using the primers P15 and P16. Both fragments were assembled and subcloned in *E. coli* Top10F' yielding plasmid VIII.24 (Table 1). Using primers P17 and P18, the region T1-P_{Rha}-RBS-*ss7036*-3xFLAG-T2-KmR-RhaS-T3 was PCR amplified with overhangs into a shortened version of pVZ322, called pVZ322s (Table 1). The backbone was digested with XmnI and both fragments were assembled and subcloned in *E. coli* Top10F´ yielding plasmid VIII.17. The plasmid VIII.18 was constructed the same way as VIII.17

1 http://parts.igem.org/Part:BBa_K2243023

10.3389/fmicb.2023.1112307

TABLE 2 Oligonucleotides used in this study. The stop codon introduced with P11 is in italics.

ID	Sequence	Description
P1	AGCTCACTCAAAGGCGGTAA	pUC19_fwd_shorterbackbone2; pos. 747 to 766fw in GenBank entry L09137.2
P2	TCACCGTCATCACCGAAACG	pUC19_rev_shorterbackbone2; pos. 25 to 6rv in GenBank entry L09137.2
P3	CGTTTCGGTGATGACGGTGAACCAAAGGAGCGGAAACC	7036-asRNA1_fwd
P4	TTACCGCCTTTGAGTGAGCTCCTAAACAGAAGGGTTTTTGC	7036-asRNA1_rev
P5	TTACCGCCTTTGAGTGAGCTCTTATAGCTCATATCCAAAACCCAG	7036-asRNA1-7037_rev
P6	TTAGAAAAACTCATCGAGCATCAAATG	pUC19_3637-Km_bb_fwd
P7	CGTTTCGGTGATGACGGTGAGAAGGGATAGCAAGCTAATTTTTATG	pUC19_pPetE_fwd
P8	CTTTGCTCATACTTCTTGGCGATTGTATC	Ppet_GFP_rev
P9	GCCAAGAAGTATGAGCAAAGG AGAAGAAC	Ppet_GFP_fwd
P10	TGCTCGATGAGTTTTTCTAATTTGTAGAGCTCATCCATG	Km_GFP_rev
P11	AGGGAGGC*TGA*TATTGCTCTG	7037_STOP_fwd
P12	GTTCAAGGTGACGGTATTG	7037_STOP_rev
P13	ATGGATTATAAAGATCATGATGG	pUC19_Rha_3xFLAG_fwd
P14	TTCTACCTCCTTTGTATATTATAAAC	pUC19_Rha_3xFLAG_rev
P15	GGAGGTAGAAATGGAAACAGGGCACTCCATCG	7036_F2_Fwd
P16	CTTTATAATCAGCGGCACCCCCTGCAAA	7036_F2_Rev
P17	GCCGACGAAGCCCGGCAGGTAAGACCC CCGCACCGAAA	RP-ilvBN_Fw
P18	CCTGCTGGGTCTGGCTCATGAATAAAAA ACGCCCGGCGGCAACCGAGCGAATTAT TGCAGAAAGCCATCCC	RhaS_oop_RP_Rev
P19	AATATACAAAGGAGGTAGAAATGAGCAAAGGAGAAGAAC	PRha_sfGFP_fwd
P20	TCATGATCTTTATAATCCATTTTGTAGAGCTCATCCATG	PRha_sfGFP_rev
P21	GGTAGAAATG*TAA*ACAGGGCAC	PRha-7036-Stop_F
P22	TCCTTTGTATATTATAAACTTACC	PRha-7036-Stop_R
P23	GCCGCCCGCATTGGAGAAATAAGACCCCCGCACCGAAA	Rp_XmnI_Prha_fwd
P24	CAGATCGTTGACGAGTATTAAATAAAAAACGCCCGGCGGCAACCG	Rp_XmnI_Prha_rev
P25	GGAAATTGCAATCTTATTCCCATTGACCC	slr7037us1000-f
P26	GGATTTATTTATTCTTACTCCACCCCCCCTGCATCA	7037us-cat-r2
P27	CAGGGCGGGGCGTAAACCTGAGTTAAACAATGGGCACC	cat-7037ds-f5
P28	CTCAATTCTAATCCCAAGCCCCCTCAG	slr7037ds1000-r
P29	CAGGGGGGGTGGAGTAAGAATAAATAAATCCTGGTG	7037us-cat-f3
P30	ATTGTTTAACTCAGGTTTACGCCCCGCCCTGCCACT	cat-7037ds-r4
P31	GCAAAATTTGCAGGGGGTGCCGCTTG	slr7037us200-f
P32	GCCATTCTTTTCCTCCATCACTGCGGTGG	slr7037ds200-r
P33	GAAGACTATCCAAATCGCCAAGG	7036 + UTR_Fwd
P34	TAATACGACTCACTATAGGG CTTAGGAACCCCTTCTGTGC	7036 + UTR_T7_Rev
P35	TAATACGACTCACTATAGGG CTAAAACCAAAGAAACCCCTC	asRNA1_T7_Fwd
P36	AGATATCAAGCGGCACCC	asRNA1_Rev
P37	CTTTAGGTGGGCGTTGACCT	antiC3S1_fwd
P38	TAATACGACTCACTATAGGGTAATAGTAATGACAGGCAG	antiC3S1_T7_rev
P39	TGGGTTCATTCAATCGACG	cmpA_fw
P40	TAATACGACTCACTATAGGGTGACGTTATCTCTAGCGGAG	cmpA_T7_rv
P41	TAATACGACTCACTATAGGGAAGACTATCCAAATCGCCAAGG	7036_InvTr_T7_Fwd
P42	TAAGACCTTAGGAACCCCTTC	7036_InvTr_Rev

(Continued)

TABLE 2 (Continued)

ID	Sequence	Description
P43	CACCTAAAACCAAAGAAACCCCTC	asRNA1_InvTr_Fwd
P44	TAATACGACTCACTATAGGGTCAAGCGGCACCCCCTGC	asRNA1_InvTr_T7_Rev
P45	CCACAATTCCAGAGGATAGCC	Primer a, Fig. 7
P46	GGGTAACTACGACCTACTCC	Primer b, Fig. 7
P47	CTGCATCAATGGCGATCGCCGACGGG	Primer c, Fig. 7
P48	TGACATGCTGAACCTGCTTGTAAAATGAG	Primer d, Fig. 7
P49	GTCGGCCCTATCTTGTTTCCGTAGC	Primer e, Fig. 7
P50	CATCAAGCCTCACAAGAGGGAGTATCC	Primer f, Fig. 7
P51	GGCAGTGATAGTACCCTGATCACCAT	Primer g, Fig. 7
P52	CACAGAACCAGACCATGGAGTGCAA	Primer h, Fig. 7
P53	AGTTGACGAAGTAGTTGTGC	Primer m, Fig. 7
P54	GAAATGGTGGCAGTGCAGG	Primer n, Fig. 7
P55	TCAAATGTTGTTCAAGGGGAAGCTGCATA	Primer o, Fig. 7
P56	CACCGACCACCACCGTAAGGCAGGGC	Primer p, Fig. 7
P57	GAAATGTTGAATACTCATACTCTTCC	Colony PCR
P58	GTATTACTGTTTATGTAAGCAGACAG	Colony PCR
P59	ATGGTTACTCACCACTGCGATCC	Colony PCR
P60	AGACGAAAGGGCCTCGTG	Colony PCR
P61	AAGTGCCACCTGACGTCTAAG	Colony PCR
P62	CGCTGTCACATTTCACAACCGAGTG	5S hybridization

but using primers P19 and P20 to amplify the sfGFP gene from plasmid VII.65.

Starting with plasmid VIII.17, the second codon of the *ssr7036* gene was modified to a stop codon (resulting in plasmid VIII.58) through site-directed mutagenesis using primers P21 and P22 and inverse PCR with plasmid VIII.24 as a template. Competent Top10F´ *E. coli* cells were transformed with the resulting PCR product. The plasmid DNA was then isolated and used as a template for PCR amplification of the mutagenized insert using primers P23 and P24. Chemocompetent Top10F' *E. coli* cells were transformed with the PCR product and XmnI digested pVZ322s plasmid as backbone.

The PCR products were assembled with AQUA cloning (Beyer et al., 2015) in chemocompetent *E. coli* DH5α cells, if not stated otherwise. The PCRBio HiFi DNA polymerase (Nippon Genetics) was used in all PCRs throughout.

To construct DNA fragments for the deletion of *slr7037*, the regions flanking *slr7037* were amplified by PCR using primers P25 and P26 (for the upstream fragment) and P27 and P28 (for the downstream fragment), respectively. The chloramphenicol resistance marker gene was amplified from pUC303 (Kuhlemeier et al., 1983) using primers P29 and P30. Three fragments were combined by PCR using primers P25 and P28. All constructs were verified by sequencing. After transformation of DNA fragments into *Synechocystis* GT-I strain (Kanesaki et al., 2012), several clones showing resistance to chloramphenicol were isolated. The complete disruption of *slr7037* in strain Δ*slr7037* was confirmed by PCR using primers P31 and P32 with genomic DNA as a template. Three clones with complete deletion of *slr7037* were selected and used for further analyses (Supplementary Figure S1).

Preparation of total RNA and Northern blot hybridizations

25 to 50 ml of densely grown cultures were harvested at 3,237 *g* for 10 min at room temperature in a swing-out centrifuge and resuspended in 1 ml PGTX (Pinto et al., 2009). For cell lysis, the samples were incubated for 15 min at 65°C with occasional vortexing, followed by addition of 700 µl chloroform/isoamyl alcohol (24:1), mixing and incubation for 10 min at room temperature. Following a second chloroform extraction, the upper phase was precipitated with 1 vol of isopropanol overnight at-20°C. Pellets were collected by centrifugation at 13,226 *g* for 30 min at 4°C and washed once with 1 ml of 70% EtOH. The air-dried pellet was dissolved in 20 to 50 µl of RNase-free H_2O.

To inactivate the temperature sensitive RNase E [*rne*(TS), (Hoffmann et al., 2021)] strains were incubated for 0 h, 1 h or 24 h at 39°C. Wild type (WT) and a strain carrying a plasmid-encoded RNase E control [*rne*(WT), resulting in higher RNase E levels than in WT] were treated the same way. For Northern blot analysis, 10 µg samples of total RNA were separated by 8 M urea 10% PAGE and electroblotted on Hybond N+ nylon membranes (Cytiva) with 1 mA per cm^2 for 1 h. Northern hybridizations were performed with radioactively labelled probes (primers P33 to P36 for templates) generated using [α-^{32}P]-UTP and the Maxiscript T7 *in vitro* transcription kit (Thermo Fisher Scientific). After blotting the RNA was crosslinked to the membrane with 240 mJ using a UV-Stratalinker (Stratagene). Hybridizations were performed in Northern buffer (50% deionized formamide, 7% SDS, 250 mM NaCl and 120 mM Na_2HPO_4/NaH_2PO_4 pH 7.2) overnight at 62°C. The membranes were washed at 57°C in buffer 1 [2 × SSC (3 M

NaCl, 0.3 M sodium citrate, pH 7.0), 1% SDS], buffer 2 (1 × SSC, 0.5% SDS) and buffer 3 (0.1 × SSC, 0.1% SDS) for 10 min each. Signals were detected with the biomolecular imager Typhoon FLA 9500 (GE Healthcare) using phosphorimaging and a photomultiplier value of 1,000 V.

Western blot analysis

Cultures were induced with 1.25 µM Cu_2SO_4 at an OD_{750} between 0.6 and 1.0. After 24 h, the cultures were collected at 3,237 *g* for 10 min, the pellet was resuspended in 200 µl TBS and transferred into a 1.5 ml screw cap tube containing 150 µl glass beads (Retsch) with diameters of 0.1 and 0.25 mm. Cell disruption was carried out with the Precellys 24 homogenizer (Bertin Technologies) at 6,000 rpm for 3 repeats of 3 × 10 s with 5 s of break in between at 4°C. After that, the mixture was transferred into a Micro Bio Spin column in an 1.5 ml Eppendorf tube and centrifuged for 2 min at 100 g at 4°C to separate the glass beads from the sample. The flow through was centrifuged at 15,871 *g* for 30 min at 4°C to remove the cell debris. The supernatant was transferred into a fresh tube and protein concentration was measured using the Invitrogen Qubit 3 fluorometer (Thermo Fisher Scientific). Ten µg total protein were loaded per lane.

5x protein loading buffer (250 mM Tris, pH 6.8, 25% glycerol, 10% SDS, 500 mM DTT, 0.05% bromophenol blue) was added to the samples before denaturation at 95°C for 5 min. As marker the PageRuler™ Prestained Protein Ladder (Thermo Fisher Scientific) was used. Samples were run on a 6% SDS polyacrylamide stacking gel and a 10% SDS polyacrylamide separating gel to separate the proteins by size. The gel was blotted onto a nitrocellulose membrane (Hybond™-ECL, Cytiva) at 1.2 mA/cm^2 for 1 h. The membrane was stained using Ponceau solution (0.2% Ponceau, 3% TCA) for 5 min shaking at 180 rpm at room temperature. To remove the Ponceau stain, the membrane was washed with deionized water. After the Ponceau staining, the membrane was washed 3 times for 5 min in TBS-T and blocked overnight in TBS-T with 3% milk powder at 4°C. The next day, the membrane was washed again 3 times for 10 min in TBS-T, then incubated for 1 h at room temperature in 20 ml TBS-T with 3% milk powder and 1:5,000 diluted anti-GFP antiserum (Abcam) while shaking with 180 rpm. Subsequently, the membrane was washed three times for 10 min in TBS-T, then incubated for 1 h in 20 ml TBS-T with 3% milk powder and 1:10,000 diluted goat anti-rabbit IgG-HRP (Sigma). Finally, the membrane was washed for 10 min twice in TBS-T and briefly once in TBS. The membrane was sprayed with WesternBright ECL Spray (Advansta) and the signal was visualized using the FUSION SL Transilluminator (Vilber Lourmat).

Total DNA preparation and Southern blot hybridization

Total DNA was prepared from *Synechocystis* 6803 by collecting cells from 50 ml cultures at an OD_{750} of 1–1.2 by centrifugation at 3,237 *g* for 10 min. The pellets were resuspended in 1 ml SET buffer (50 mM Tris, pH 7.5; 1 mM EDTA, 25% (w/v) sucrose). For lysis, resuspended cells were incubated in 100 mM EDTA (pH 8), 2% SDS (w/v) and 100 µg/ml proteinase K at 50°C for 16 h. DNA was extracted by phenol/chloroform/isoamylalcohol (25:24:1) extraction. DNA was precipitated from the final aqueous phase by adding 1 vol isopropanol, incubation at-20°C for at least 2 h and centrifugation at 15,871 *g* for 30 min at 4°C. The pellet was washed with 70% ethanol, air-dried and resuspended in 30 µl sterile Milli-Q water.

For restriction analysis, 4 µg of total DNA each were digested with FastDigest HindIII restriction endonuclease for 3 h at 37°C to ensure complete digestion. The digested DNA samples were subjected to gelelectrophoretic separation for 1 to 2 h at room temperature on ethidium bromide stained 0.8% agarose gels with 120 V. Thereafter, the gels were incubated at 70 rpm at room temperature for 30 min each in 0.25 M HCl, in denaturation solution (1.5 M NaCl, 0.5 M NaOH) and in neutralization solution (1.5 M NaCl, 0.5 M Tris pH 7.5) prior to blotting. The DNA samples were blotted onto Hybond-N+ nylon membranes (Cytiva) by capillary transfer overnight with 20x SSC (3 M NaCl, 300 mM sodium citrate pH 7) as transfer buffer. After blotting the DNA was crosslinked to the membrane with 120 mJ using a UV-Stratalinker (Stratagene).

For the generation of isotope-labelled probes, templates were amplified *via* PCR, using the primers P38 and P40 containing the T7 promoter sequence together with the respective primers P37 and P39 and genomic DNA from *Synechocystis* 6803. Transcript probes labelled with [α-^{32}P]-UTP (3,000 Ci/mmol, 10 mCi/ml) were generated from these templates using the MAXIscript® T7 *In Vitro* Transcription Kit (Thermo Fisher Scientific).

Hybridization was performed overnight at 52°C in hybridization buffer followed by 10 min wash steps each in wash buffers 1, 2 and 3 (for buffers see Northern blot section) at 47°C. The signals were detected with a Phosphor Imaging Screen (Bio-Rad) and the GE Healthcare Typhoon FLA 9500 imaging system.

Endoribonuclease assays

Purification of 6xHis-tagged RNase E was performed as published (Behler et al., 2018). In short, codon-optimized and TEV site-fused *slr1129* from *Synechocystis* 6803 was expressed under control of an IPTG-inducible promoter on pQE70. Expression was induced with 1 mM IPTG in *E. coli* M15 [pREP4] at an OD_{600} of 0.7 at 22°C for 24 h. The cells were collected by centrifugation and the recombinant protein was column-purified using Ni^{2+}-NTA resin (Qiagen). Elution fraction 2 was used for subsequent cleavage assays. The purified protein was stored at-80°C. The protein concentration was measured using the Qubit 3.0 fluorometer and Qubit Protein Assay Kit (both from Thermo Fisher Scientific). RNase E was used at a final concentration of 0.044 µM (2%) - 2.2 µM (100%).

In vitro transcription of *ssr7036* (consisting of the complete 5' UTR and the first 100 nt of coding sequence) and asRNA1 substrates for RNase E assays was performed with the MEGAshortscript T7 transcription kit (Thermo Fisher Scientific). Suitable templates were PCR-amplified (primers P41 to P44) and thereby tagged with a T7 promoter. The amplified and purified fragments were used for *in vitro* transcription according to the manufacturer's specifications. *In vitro* transcripts were gel purified using 8 M urea 10% PAA and the ZR small-RNA PAGE Recovery Kit (Zymo Research). The *in vitro* transcribed RNAs carry three additional G nucleotides at their 5′ ends originating from the T7 promoter. To generate suitable substrates for RNase E, the transcripts were dephosphorylated by RNA polyphosphatase (RPP, Lucigen) generating a 5' monophosphate. For RPP treatment 5 µg gel purified RNA, 20 U RPP, 2 µl RPP buffer and 0.5 µl RiboLock (Thermo Fisher Scientific) were incubated in a total volume of 20 µl for 30 min at 37°C and

subsequently column-purified using RNA Clean and concentrator 5 (Zymo Research).

Cleavage reactions were performed in a volume of 10 µl in RNase E cleavage buffer (Behler et al., 2018). Transcripts (0.4 µM) and buffer were premixed, incubated for 5 min at 65°C, cooled down to room temperature, then 2 µl of RNase E or dilutions as indicated were added. As negative control, 2 µl elution buffer 1 were added. Reactions were incubated for 15 min at 30°C, stopped by the addition of 2× RNA loading dye (95% formamide, 0.025% SDS, 0.025% bromophenol blue, 0.025% xylene cyanol FF, 0.5 mM EDTA, pH 8). Half of the reactions were loaded onto denaturing 8 M urea 10% PAA gels. RiboRuler low range RNA ladder (Thermo Fisher Scientific) and low range ssRNA ladder (New England Biolabs) were used as size markers. RNA was visualized with SYBR® Gold Nucleic Acid Gel Stain 1:10,000 diluted in 0.5x TBE. Signals were detected with the biomolecular imager Typhoon FLA 9500 (GE Healthcare). Excitation of 473 nm, emission filter long pass blue ≥510 nm and a photomultiplier value of 400–550 V were used.

Analysis of the genomic structure of *slr7037* knockout strains

Genomic DNA was extracted from cyanobacterial cells as described (Kanesaki et al., 2012). The purity of the genomic DNA extract was checked in an Agilent 2,100 bioanalyzer (Agilent Technologies, Palo Alto, Calif.). Since the *Synechocystis* 6803 genome contains many duplicated sequences exceeding 1 kb, such as transposons, we selected a mate-pair sequencing approach, suitable for longer DNA fragments (2–10 kb) to analyse the genomic structure. Sequencing libraries were prepared using the Nextera Mate Pair Library Preparation Kit (Illumina). Mate-pair sequencing was carried out for 150 cycles using the Nextseq550 system (Illumina Inc., CA, United States) according to the manufacturer's specifications. Original sequence reads were deposited in the DRA/SRA database with accession numbers DRX398828 to DRX398831.

The sequencing reads were trimmed using the CLC Genomics Workbench ver. 21.0.3 (Qiagen, Venlo, Netherland) with the following parameters: Phred quality score >30; removing the terminal 5 nt from both 5′ and 3′ ends; and removing truncated reads <20 nt. The junction adapter sequences originating from the library construction were removed to avoid obstructions to read mapping according to the instructions. 4,000,000 trimmed reads were randomly extracted and mapped to the reference genome sequence of *Synechocystis* 6803 (accession numbers: NC_020286.1, NC_020287.1, NC_020288.1, NC_020289.1, NC_020290.1, NC_020296.1, NC_020297.1, NC_020298.1) using CLC Genomics Workbench ver. 21.0.3 (QIAGEN) with the following parameters: length fraction: 0.8, similarity fraction: 0.9.

The procedure for investigating the integration of pSYSA into the chromosome or other plasmids is shown in Supplementary Figure S2. First, pairs were selected in which one side of the paired reads mapped to pSYSA. Then, this set was narrowed down to pairs in which the other side mapped to a chromosomal or plasmid locus. The same procedure was performed by switching references, and read pairs common to both subsets were extracted. The first position of each read is plotted in Supplementary Figure S2.

The average read depth was calculated by dividing the number of reads mapped to a specific replicon by the length of this replicon. The ratio of read depths of plasmids to chromosome were used as measure for the plasmid copy number relative to chromosomal copies (Supplementary Table S1).

Results

The most abundantly transcribed region on *Synechocystis* 6803 plasmid pSYSA consists of an mRNA:antisense RNA locus

We observed previously that RNase E expression level affected pSYSA plasmid copy number (Hoffmann et al., 2021). Because overlapping transcripts play a central role in the copy number control of various plasmid replication systems (Lin-Chao and Cohen, 1991; Brantl et al., 1993; Hiraga et al., 1994; Malmgren et al., 1996, 1997), we searched for a pair of overlapping abundant transcripts on the plasmid pSYSA. The most abundantly transcribed region beyond the three CRISPR systems on pSYSA is indeed a locus transcribed from both strands. It contains the short protein-coding gene *ssr7036* on the forward strand and a partially overlapping asRNA, here called asRNA1, transcribed from the reverse strand (Figure 1). The *ssr7036* mRNA originates from a single transcriptional start site (TSS) at position 32,177 on the forward strand. The previous classification of transcripts assigned the mRNA to two separate transcriptional units (TUs), TU7029 and TU7030 (Kopf et al., 2014), with TU7030 encompassing the coding sequence of *ssr7036* and beginning just 5 nt before the start codon. Previously, putative processing sites (PSS) in *Synechocystis* 6803 were defined transcriptome-wide using TIER-seq (Hoffmann et al., 2021). Here, our reanalysis of data on PSS and TSS from that study (Hoffmann et al., 2021) revealed that TU7029 and TU7030 derive from a single precursor transcript upon processing by RNase E.

The asRNA1 transcript accumulates to a very substantial level. It originates from a single TSS at position 32,481 on the reverse strand and starts with an adenine which is complementary to the thymidine that is the first nucleotide of the *ssr7036* TGA stop codon. The length of asRNA1 according to differential RNA-seq (dRNA-seq) data (Kopf et al., 2014) is 153 nt (Figure 1A). Over its entire length, asRNA1 is complementary to the coding sequence of *ssr7036* (Figure 1B). In dRNA-seq data, the asRNA1 level was lowest in the dark, but accumulated to similar levels under all other conditions. In contrast, TU7029 showed widely changing levels under different conditions. The dRNA-seq data are focused on the primary transcript's 5′ ends, which result from transcriptional initiation (Mitschke et al., 2011). Hence, full-length transcripts can accumulate differently than suggested by dRNA-seq. To control for this, we performed Northern blot hybridizations to judge the actual accumulation of asRNA1 and *ssr7036* mRNA. For asRNA1, a minor and a major transcript accumulated slightly above and below 150 nt (Figure 1C). This is consistent with the observed dRNA-seq coverage and indicates some heterogeneity at the 3′ end.

For *ssr7036*, signal intensities differed strongly between different conditions. Signals were very low or undetectable for the samples from stationary phase, heat shock at 42°C or darkness, but prominent in samples from exponential growth phase and several other conditions such as cold shock at 18°C, high light or low iron (Figure 1C). A particularly long transcript of ~350 nt was observed in the samples from stationary phase, cold shock, high light and phosphate depletion (upper black arrowhead in Figure 1C). Moreover, multiple bands with sizes of ~140 nt and in the range of 200 to 260 nt (Figure 1C) were observed. These multiple bands likely result from processing, because no

A

TU ID	TU start	TU end	Strand
TU7029	32177	32283	+
TU7030	32284	32632	+
TU7031	32481	32329	-

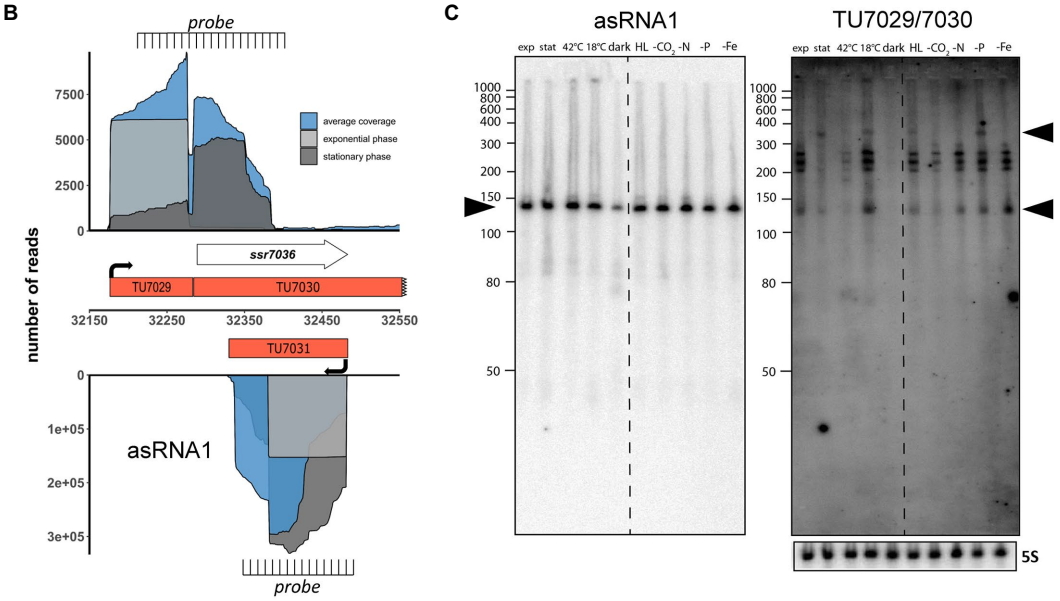

FIGURE 1
The *ssr7036*/asRNA1 locus on plasmid pSYSA in *Synechocystis* 6803. **(A)** Definition of the transcriptional units TU7029, TU7030 and TU7031 according to previous analyses by dRNA-seq (Kopf et al., 2014). The TU start and end positions are given for the *Synechocystis* 6803 pSYSA sequence available in GenBank under accession number AP004311. **(B)** Location of TUs according to the previous annotation of the transcriptome and genome-wide mapping of transcriptional start sites (Kopf et al., 2014), together with the mapping data from previous dRNA-seq analyses for two conditions, exponential (light gray) and stationary (dark gray) growth phase and the average total coverage (blue). The locations of two probes used in panel C for Northern blot analyses are indicated. Please note that the TU7030 continues beyond the end of the *ssr7036* gene, although with lower coverage. **(C)** Northern blot hybridization of total RNA using a ^{32}P-labeled transcript probe specific for asRNA1 (TU7031, left) and *ssr7036* (TU7029/TU7030, right panel) subsequently on the same membrane after separation of 12µg total RNA isolated from cultures grown under ten different growth conditions on a denaturing 10% polyacrylamide gel. The different growth conditions (Kopf et al., 2014) were: exponential phase (exp.), stationary phase (stat.), heat stress for 30min (42°C), cold stress for 30min (15°C), darkness for 12h (dark), high light, 470µmol photons $m^{-2}s^{-1}$ for 30min (HL), depletion for inorganic carbon, cells were washed 3 times with carbon-free BG-11 and cultivated further for 20h (-C), nitrogen starvation, cells were washed 3 times with nitrogen-free BG-11 and cultivated further for 12h (-N), phosphorus starvation, cells were washed 3 times with phosphorus-free BG-11 and cultivated further for 12h (-P), iron depletion by adding the iron-specific chelator desferrioxamine B and continued cultivation for 24h (-Fe). The locations of the two probes are indicated in panel B. The asRNA1 and *ssr7036* transcripts accumulate differently under the various growth conditions. Black arrowheads point at the major accumulating transcript form for asRNA1 and in case of *ssr7036,* two transcripts that match approximately the expected lengths for TU7030 and TU7029. A hybridization for the 5S rRNA using the labeled oligonucleotide P62 (Hein et al., 2013) was used for loading control. RiboRuler low range RNA ladder (Thermo Fisher Scientific) and low range ssRNA ladder (New England Biolabs) were used as size markers.

additional TSS was mapped to this region. One should note that the used probe is complementary to both TU7029 and TU7030.

We conclude that the *ssr7036* mRNA and asRNA1 form a pair of overlapping transcripts originating from pSYSA. The less abundant *ssr7036* mRNA accumulates to different levels in various environmental conditions and appears to be processed at multiple positions out of a precursor transcript spanning TU7029 and TU7030.

The level of RNase E expression impacts *ssr7036* transcript accumulation

To study possible effects of changing RNase E availability *in vivo*, we used the previously generated strains *rne*(WT) and *rne*(Ts) (Hoffmann et al., 2021). These strains were constructed by introducing the complete *rne-rnhB* locus, including the native promoter and 3' UTR, on a self-replicating plasmid into *Synechocystis* 6803 WT followed by deletion of the *rne-rnhB* genomic locus by homologous recombination (Hoffmann et al., 2021). The strain in which the native, unchanged *rne* gene was introduced was called *rne*(WT). The temperature-sensitive strain *rne*(Ts) was constructed in the same manner, but contains mutations leading to the two amino acid substitutions I65F and V94A in the *rne* gene. A temperature of 39°C was established as leading to RNase E inactivation and causing substantial physiological side effects over time, which were not observed at the permissive temperature of 30°C (Hoffmann et al., 2021). For these strains, we expected the following effects:

- Transcripts accumulating with increasing time at 39°C at a higher level in *rne*(Ts) compared to *rne*(WT) are likely direct targets of RNase E
- Transcripts with a decreasing level in *rne*(Ts) over time at 39°C indicate RNA species which likely require the enzyme for maturation from a less stable precursor

- Changed transcript abundance in *rne*(WT) compared to *Synechocystis* 6803 WT result from the higher gene dosage of the plasmid-located *rne-rnhB* operon compared to the chromosomal locus.

We grew cultures of *Synechocystis* 6803 WT, *rne*(WT) and *rne*(Ts) in quadruplicates and subjected them to the non-permissive temperature of 39°C. Total RNA was prepared immediately before the temperature shift (0 h), 1 h and 24 h after. The *ssr7036* transcript and asRNA1 levels were determined by Northern blot hybridization. At 0 h, we observed an increased *ssr7036* signal intensity in *rne*(WT) and *rne*(Ts) compared to *Synechocystis* 6803 WT (Figure 2A). This is most likely explained by elevated RNase E levels in *rne*(WT) and *rne*(Ts) compared to WT due to the higher *rne* gene dosage. After 1 h and 24 h at 39°C, the transcript pattern in *rne*(Ts) was markedly changed. A new band appeared in *rne*(Ts) at 1 h while the strongest signals in *Synechocystis* 6803 WT and *rne*(WT) became weakened in *rne*(Ts) at 24 h after the temperature upshift.

For the asRNA1 clear differences were observed in *rne*(WT) and WT compared to *rne*(Ts) at 1 h and 24 h after the temperature upshift. Transcripts just below the 50 nt marker band and in the region between 80 and ~90 nt were present in *rne*(WT) and WT, barely detectable in *rne*(Ts) at 0 h and disappeared in *rne*(Ts) with increasing time at the non-permissive temperature (Figure 2B). Therefore, these asRNA1-derived transcripts depended on RNase E activity.

We conclude that the *in vivo* accumulation of different *ssr7036*-related transcript types in sense as well as antisense orientation changed upon RNase E inactivation.

Mapping of RNase E cleavage sites *in vivo*

To precisely map the RNase E cleavage sites *in vivo*, we reanalyzed the previously prepared transcriptome-wide datasets of RNase E cleavage sites. After transient inactivation of RNase E (TIER-seq) in the temperature-sensitive strain, *rne*(Ts) (Hoffmann et al., 2021), processing sites with higher read counts in *rne*(WT) compared to *rne*(Ts) indicated RNase E cleavage sites. We identified several RNase-E-dependent processing sites within TU7029, TU7030 and asRNA1 (Figure 3A; Table 3).

In the second analyzed dataset, a strain harbouring a 5′-sensing-deficient RNase E variant, *rne* (5p), was compared to *rne*(WT) and WT. 5′ sensing describes a central mechanism by which RNase E can recognize its targets (Hoffmann et al., 2023). In this analysis, processing sites with higher read counts in *rne*(WT) or *rne* (5p) compared to *Synechocystis* WT indicated cleavage sites of the RNase E native enzyme or the 5′-sensing-deficient enzyme, respectively. Again, several RNase-E-dependent processing sites were identified (Figure 3B; Table 3).

Altogether, ten PSS were identified within TU7029 and TU7030. Three of these were located at adjacent nucleotide positions and showed a similar response to the manipulation of RNase E activity. The cleavage sites located at position 32,284/32,285, at the border between TU7029 and TU7030, and at position 32,293/32,294, 4/5 nt into the *ssr7036* coding sequence (the reading frame including stop codon extends from position 32,289 to 32,483) (Figure 3A) were significantly more prominent in *rne*(WT) compared to *rne*(Ts) (Table 3). Therefore, these PSS represent major RNase E processing sites. It is noteworthy that these sites correspond to a pronounced shift in the read coverage. While TU7029 accumulated to a slightly lower level in *rne*(WT) compared to *rne*(Ts), this difference was inversed at the position of these PSS, and TU7030 read coverage was slightly higher in *rne*(WT) than in *rne*(Ts) (Figure 3A, upper panel). We conclude that processing by RNase E contributes to a stabilization of TU7029 and a destabilization of TU7030.

Further peaks in TU7029 and TU7030 accumulated at positions 32,250/32,251 and 32,260 in *rne* (5p) compared to *rne*(WT), indicating that a further processing of the respective RNA fragments is dependent on 5′ sensing (Figure 3B). No further strong processing sites were detectable within the *ssr7036* coding sequence, except the sites 4/5 nt into the coding sequence. This is likely due to duplex formation with asRNA1, shielding this region from attacks by the single-strand-specific RNase E.

Four processing sites were detected in asRNA1. We noticed that the cleavage site at 32,463/32,464, 18/19 nt downstream of the TSS of asRNA1, was more prominent in *rne*(Ts) than in *rne*(WT) (Table 3). Hence, the respective RNA fragments were stabilized due to the declining RNase E activity in *rne*(Ts) after 1 h at the non-permissive temperature. This indicates that this cleavage likely is performed by another RNase, but that the resulting RNA fragment would normally be further processed by RNase E. To the contrary, another two cleavage sites, located at 32,426 and 32,418, accumulated significantly more strongly in *rne*(WT) compared to *rne*(Ts), indicating these were *bona fide* RNase E PSS (Table 3). The mapped cleavage sites correspond to several of the fragments observed in the Northern blots (Figures 1C, 2).

In vitro analysis of RNase E cleavage sites within the *ssr7036* precursor and asRNA1

To confirm direct RNase E cleavages within the *ssr7036* precursor and asRNA1, *in vitro* cleavage assays were performed. According to the dRNA-seq data set, some asRNA1 transcripts started at the positions just upstream of the main TSS at position 32,481 (Figure 1A; Table 3). Hence, we included the two preceding nucleotides (positions 32,482 and 32,483 on the reverse strand) in the 153 nt long full-length asRNA1, which was used as a substrate. For the *ssr7036* precursor transcript, we selected a substrate encompassing the entire TU7029 and the first 105 nt of TU7030. Additionally, both substrates contained three additional Gs at their 5′ ends left from the initiation of transcription at the T7 RNA polymerase promoter, yielding transcripts of 158 nt (asRNA1) and 218 nt (*ssr7036*).

The incubation with recombinant RNase E yielded three major fragments for asRNA1 that were not further degraded when increasing the amount of enzyme (Figure 4A). Their sizes match the expected lengths of 98, 90 and 60 nt, in case cleavage occurred at sites corresponding to positions 32,426 (yielding the 60 and the 98 nt fragments) and 32,418. The latter cleavage would yield the 90 nt fragment by cleaving off 8 nt from the 98 nt fragment. This interpretation is consistent with the *in vivo* cleavage site mapping, because the processing sites at 32,426 and 32,418 declined in the temperature-sensitive mutant after 1 h at 39°C (Figure 3). In contrast, the other site at 32,463/32,464 was likely caused by a different enzyme *in vivo*, as we did not observe fragments matching these positions in the *in vitro* assays (Figure 4A).

RNase E processed the *ssr7036* precursor transcript into several lowly abundant fragments and one prominent fragment of approximately 70 nt (Figure 4A). When both *in vitro* transcripts were mixed in a 1:1 ratio and incubated with RNase E, larger cleavage products were observed than in the single digests. This indicates the formation of stable

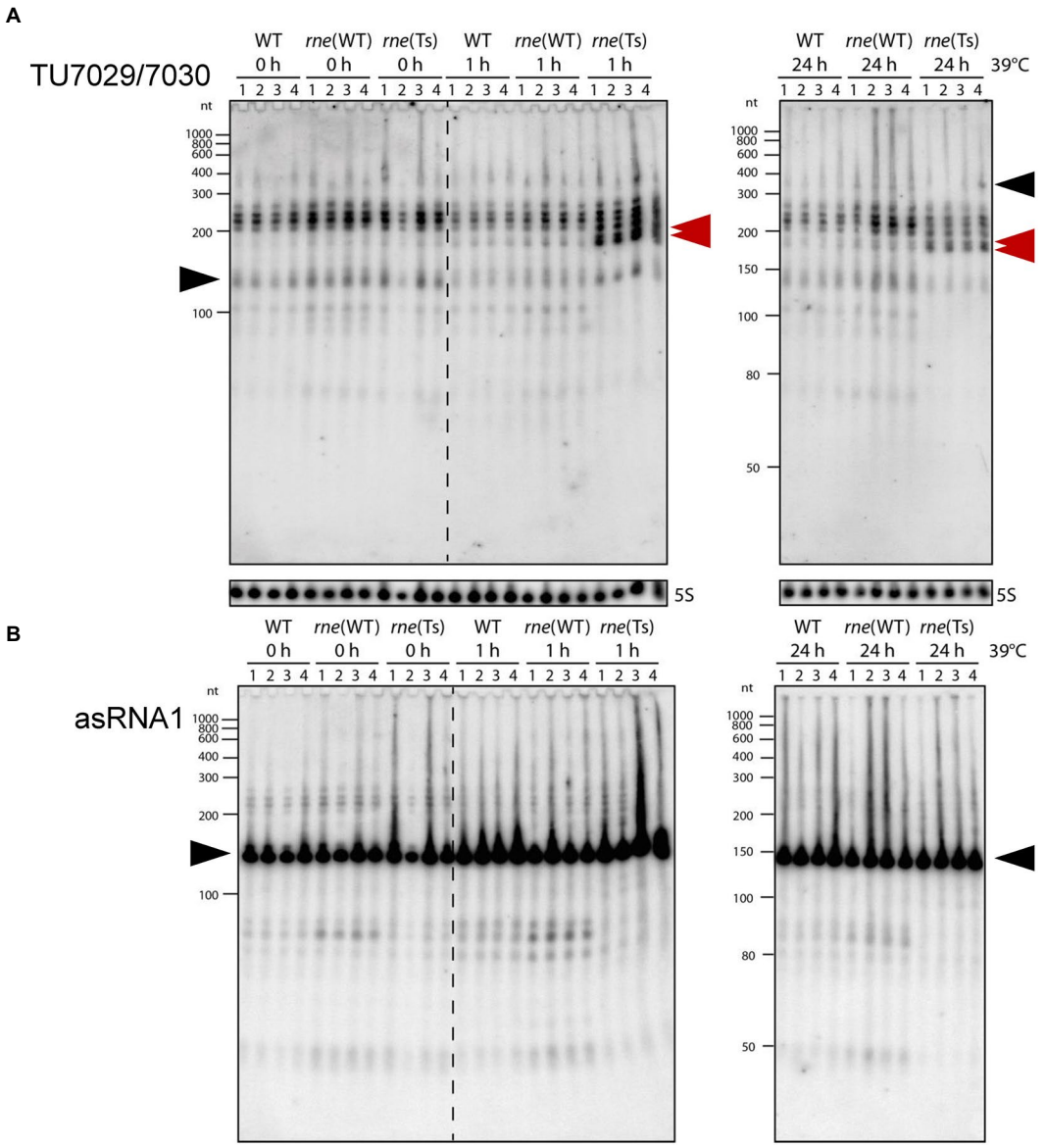

FIGURE 2
Accumulation of *ssr7036* and asRNA1 in *Synechocystis* 6803 WT *rne*(WT) and a temperature sensitive RNase E variant *rne*(Ts). Both strains contain an additional plasmid-encoded RNase E gene, resulting in higher RNase E levels compared to WT. Strains were exposed to 39°C to heat-inactivate RNase E in the strain *rne*(Ts) for 0h, 1h, or 24h (four biological replicates). Ten µg total RNA were size-fractionated by denaturing 8M urea 10% PAGE. **(A)** Hybridization against the TU7029/7030 encompassing *ssr7036*. A double band showing up in *rne*(Ts) at 1h and 24h is labeled by two red arrowheads. The same probes were used as in Figure 1C. **(B)** Hybridization against asRNA1. Black arrows as in Figure 1C. Ten µg total RNA were separated per lane and a 5S rRNA hybridization was performed to show equal loading. As molecular mass markers, RiboRuler low range RNA ladder (Thermo Fisher Scientific) and low range ssRNA ladder (New England Biolabs) were used.

RNA duplexes between both molecules. While the band pattern differed quite strongly between the cleavage assays of individual substrates and their mixture, the major 70 nt fragment of the *ssr7036* precursor transcript and the 60 nt asRNA1 fragment did not change. This suggests that these two fragments did not form duplexes in this assay.

The overlapping region between the two substrates was 62 nt (Figure 4B). Thus, other signals which did not appear in the cleavage assay with single substrates were also protected by RNA–RNA interactions. Likely, RNA secondary structures and long-range interactions have contributed to the observed cleavage pattern as well.

In summary, we could verify several RNase E cleavage sites in both transcripts *in vitro*. The resulting fragments involved the coding region of *ssr7036* and asRNA1 and originated from regions which were not protected by RNA:RNA interactions. The cleavage site at 32250/32251 and the two RNase E sites flanking the start codon of *ssr7036* 32284/32285 and 32293/32294 in Table 3 were not protected by interaction with the asRNA1.

Expression of *ssr7036* in *trans* leads to a higher ratio of pSYSA to chromosome

To study a possible effect of *ssr7036* on the pSYSA copy number, we engineered plasmid VIII.17 on basis of the vector pVZ322s in which

10.3389/fmicb.2023.1112307

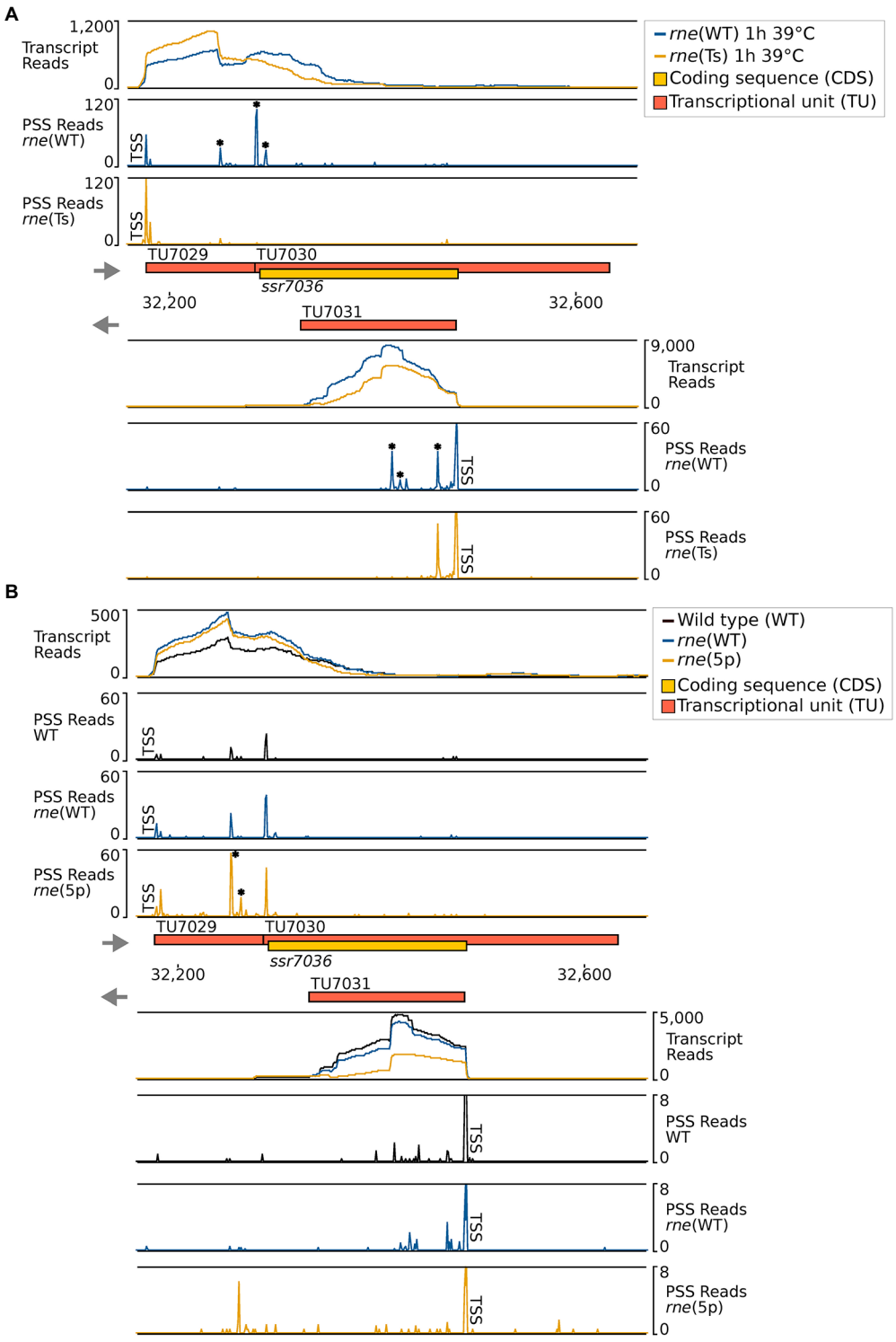

FIGURE 3
RNase-E-dependent processing events in the *ssr7036*/asRNA1 locus (TU7029, TU7030 and TU7031). **(A)** TIER-seq analysis comparing a strain encoding a temperature-sensitive RNase E variant, *rne*(Ts), to one expressing wild-type RNase E, *rne*(WT), after a heat shock of 1h at 39°C. This heat shock inactivates RNase E in *rne*(Ts) and enables the identification of RNase-E-dependent cleavage sites (Hoffmann et al., 2021). **(B)** Comparison of processing sites in wild-type *Synechocystis* 6803 (WT), in strain *rne*(WT) and a strain harbouring 5′-sensing-deficient RNase E, *rne* (5p). For both panels A and B, transcriptome coverage is given on top for the two, respectively three indicated strains. Cleavage sites are displayed in diagrams below by black, blue and orange peaks, representing 5′-monophosphorylated RNA ends (processing sites, PSS) detected in the different strains. 5′-triphosphorylated RNA ends (transcriptional start sites, TSS) may be converted to 5′-monophosphorylated RNA ends *in vivo* or during RNA-seq library preparation. Thus, TSS are partially detected in the PSS signal. Positions which were classified as TSS are indicated by "TSS" next to the respective peaks. Data was analyzed as described previously (Hoffmann et al., 2021, 2023) using DESeq2 (Love et al., 2014). DESeq2 assigns adjusted *p* values based on the Wald test and uses the Benjamini-Hochberg procedure to correct for multiple testing. Peaks with statistically significantly different read counts between the different strains were indicated by small asterisks next to them (*). Transcriptome coverage and cleavage sites (PSS) represent the average of normalised read counts of the investigated replicates [3 each for *rne*(WT), *rne*(Ts) and WT, 2 for *rne*(5p)].

TABLE 3 TIER-seq mapped RNase E cleavage sites in the transcriptional units encompassing *ssr7036* and asRNA1.

				rne(WT/Ts)		*rne*(WT/5p)	
P	S	TU	Comments	FC	p.adj	FC	p.adj
32,181	+	TU7029	4 nt downstream of TSS of TU7029	−2.40	0.00	−2.13	0.11
32,250	+	TU7029	smaller peak in TU7029	1.13	0.14	−1.34	0.09
32,251	+	TU7029	smaller peak in TU7029	0.87	0.41	−2.17	0.01
32,260	+	TU7029	showing up in rne(5p) data set	2.06	0.52	−3.56	0.03
32,284	+	TU7030	defines TU border; 5 nt upstream of *ssr7036* start codon, major peak	5.85	0.00	2.09	0.04
32,285	+	TU7030	4 nt upstream of *ssr7036* start codon	4.28	0.00	−0.12	0.94
32,293	+	TU7030	4 nt into *ssr7036* CDS	3.62	0.14	0.91	NA
32,294	+	TU7030	5 nt into *ssr7036* CDS	5.72	0.01	0.68	0.82
32,469	+	TU7030	small peak	1.33	0.42	−1.56	NA
32,472	+	TU7030	small peak	−1.61	0.07	−0.40	NA
32,464	–	TU7031	big peak in asRNA1	−1.06	0.18	0.00	NA
32,463	–	TU7031	big peak in asRNA1	−1.02	0.03	1.35	NA
32,426#	–	TU7031	smaller peak, absent in *rne*(Ts)	3.97	0.02	3.70	NA
32,418#	–	TU7031	big peak absent in *rne*(Ts)	4.16	0.00	−0.59	NA

The nt position (P) and strand (S) of the respective site is given on pSYSA, followed by the TU ID and comments. In the last four columns, the comparison for the differences in peak heights are given for *rne*(WT) versus *rne*(Ts) and *rne*(WT) versus *rne*(5p), expressed as $\log_2$FC (FC) and whether these were significant (if p.adj < 0.05, shaded). Hash symbols (#) indicate whether a site matches the results of *in vitro* cleavage analysis in Figure 4.

ssr7036 was ectopically expressed under control of a rhamnose-inducible promoter (P_{rha}). In this construct, the asRNA1 promoter elements are not included and hence asRNA1 is not transcribed from the plasmid. In parallel, a second plasmid, VIII.58, was constructed in which the second codon in *ssr7036* (GAA encoding glutamate) was converted into an UAA stop codon. Except of this single nucleotide difference, the plasmids VIII.17 and VIII.58 were identical.

Southern blot analyses were performed with three independent clones each of the *ssr7036* overexpressor (VIII.17) and the VIII.58 strain containing the *ssr7036* nonsense mutation on the plasmid (VIII.58). As controls we used three technical replicates of the WT and three biological replicates each of strains carrying either the empty vector pVZ322s (plasmid VI.16) or a plasmid in which a *sfGFP* gene was inserted instead of *ssr7036* (VIII.18; see Table 1 for an overview of all constructs). Prior to DNA preparation, all cultures were divided into two aliquots. One of these aliquots was cultivated in the presence of 3.3 mM rhamnose for 24 h to trigger transcription from the P_{rha} promoter while the other aliquot was not. To analyze the copy number of the pSYSA plasmid relative to the number of chromosomal copies, Southern blot hybridization was performed using two probes. One probe was directed against the unrelated CRISPR3 locus on pSYSA (Scholz et al., 2013), while the other was directed against the *cmpA* locus on the chromosome (gene *slr0040* encoding a bicarbonate binding protein).

The single bands observed for either probe are consistent with the expected HindIII restriction fragment sizes (9.6 kb for *cmpA*; 2.55 kb for CRISPR3) (Figure 5A). Moreover, both mutant strains, VIII.17 as well as VIII.58, showed a stronger signal for pSYSA (CRISPR3 probe) relative to the chromosomal signal (*cmpA* probe) compared to the WT and the two other controls. This change in signal intensity was not caused by the presence of the additional plasmid in the cells, because the VI.16 and VIII.18 controls contained the same vector backbone.

Next, we quantified the obtained differences in signal intensities by dividing the averaged signal intensities for the CRISPR3 fragments by the averaged signal intensities for the *cmpA* fragments. The pSYSA copy number was significantly increased by factors of 2.75 and 3.2 in the VIII.17 and VIII.58 mutant strains compared to the WT (Figure 5B). The signal intensities between the VIII.17 and VIII.58 strains were not significantly different, indicating that translation of *ssr7036* was not required for the higher pSYSA copy number. It seems that *ssr7036* transcription was sufficient for this effect. This result supports our idea that the ratio between the *ssr7036* and the asRNA1 transcript amount and their RNase E-mediated cleavage and degradation is important for copy number control of the plasmid pSYSA. In a parallel study, a 3.8-fold increase in the pSYSA copy number was observed in strains overexpressing RNase E and RNase HII (Hoffmann et al., 2023), which matches the results of this work. Hence, we conclude that an elevated *ssr7036* transcript level leads to a higher pSYSA copy number.

The *ssr7036*/asRNA1 locus together with gene *slr7037* is required for extrachromosomal replication

In the preceding section, we connected the elevated *ssr7036* transcript level to a higher pSYSA copy number. Therefore, we wanted to test if the *ssr7036*/asRNA1 locus could function as an origin of replication. For this purpose, we inserted the *ssr7036*/asRNA1 locus together with the native *ssr7036* 5′ and 3’ UTRs into the plasmid pUC19s yielding plasmid VIII.22 (Table 1). The pUC19s origin of replication is

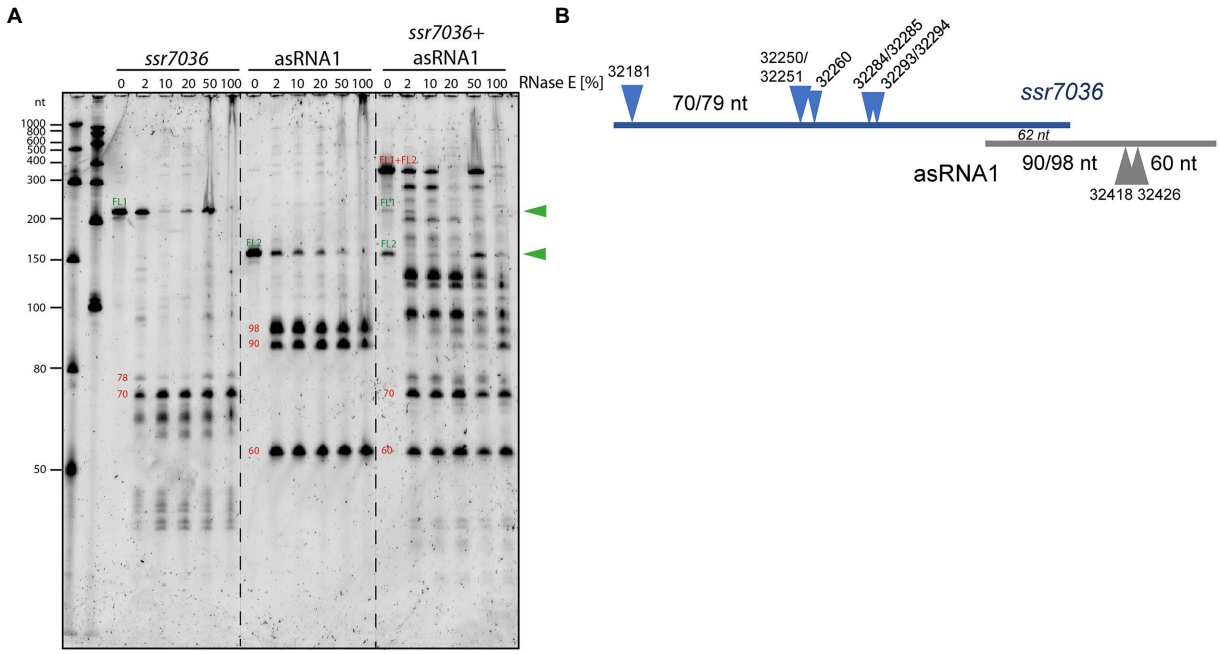

FIGURE 4
In vitro cleavage assays with purified RNase E and potential RNA substrates. **(A)** 5′ monophosphorylated *in vitro* transcripts of *ssr7036*, asRNA1 or both were mixed with cleavage buffer, incubated for 5min at 65°C, cooled down at room temperature and then mixed with 2μL of recombinant RNase E. Different dilutions of RNase E were used ranging from 1:50 (= 2%) up to undiluted (100%), as indicated. Reactions were incubated for 15min at 30°C. Half of the reactions were size-fractionated by denaturing 8M urea 10% PAGE. RiboRuler low range RNA ladder (Thermo Fisher Scientific) and low range ssRNA ladder (New England Biolabs) were used as size markers. RNA was visualized with SYBR® Gold Nucleic Acid Gel Stain. The full-length transcripts (FL1, FL2) are indicated by green arrowheads. The sizes of putative cleavage fragments are given in red numbers. **(B)** Scheme for the cleavage sites and major accumulating fragments in panel A or mentioned in the text.

not functional in cyanobacteria. Hence, plasmid pUC19s does not replicate in this group of bacteria. Accordingly, transformation of *Synechocystis* 6803 (Supplementary Figure S3A) as well as *Synechococcus* 7942 (Supplementary Figure S3B) was hardly possible with the VIII.22 plasmid, yielding only very few or no colonies, respectively. The experiments were repeated three times, including three technical replicates. We conclude that the *ssr7036*/asRNA1 locus did not function as origin of replication for the modified pUC19s plasmid in *Synechococcus* 7942 and only very rarely in *Synechocystis* 6803. A possible explanation for the observed few colonies for *Synechocystis* 6803 transformation could be a factor present in this organism which is lacking in *Synechococcus* 7942. Indeed, the *ssr7036*/asRNA1 locus is next to the *slr7037* gene, which encodes a 958 amino acid protein with predicted primase and helicase domains (Supplementary Figure S4A), suggesting it might be involved in pSYSA replication. Therefore, we constructed another pUC19s derivative in which the entire contiguous fragment was included, stretching from the *ssr7036*/asRNA1 locus to the end of *slr7037*. The resulting plasmid, VIII.23 (Table 1), was used for transformation of *Synechocystis* 6803 and *Synechococcus* 7942 and now colonies were observed for both cyanobacteria (Supplementary Figure S3). We conclude that the entire fragment encompassing the *ssr7036*/asRNA1 locus and the downstream located gene *slr7037* was needed to achieve plasmid replication. However, the numbers of colonies were consistently one to two orders of magnitude higher in *Synechococcus* 7942 than in *Synechocystis* 6803, which likely was due to competition with the native pSYSA plasmid present in the latter. Moreover, we introduced a nonsense mutation into *slr7037* by converting codon 64 into an opal stop codon (Supplementary Figure S4B) yielding plasmid VIII.45. Transformation of VIII.45 yielded no colonies for *Synechococcus* 7942 and a very low number of colonies for *Synechocystis* 6803 (Supplementary Figure S4C). Again, this result can be explained by the pSYSA-encoded native Slr7037 protein in *Synechocystis* 6803 acting in trans.

To test if the plasmid VIII.23 can function as a vector for the expression of cargo genes, we inserted a cassette consisting of the copper inducible *petE* promoter (P_{petE}) (Zhang et al., 1992) and the *sfGFP* gene encoding the superfolder green fluorescent protein as a reporter, yielding plasmid VIII.44 (Table 1). When *Synechocystis* 6803 and *Synechococcus* 7942 were transformed with the plasmid VIII.44, colonies were observed for both strains (Figure 6), confirming that the plasmid replicates in these cyanobacteria. Again, the numbers of colonies were one to two orders of magnitudes higher in *Synechococcus* 7942 than in *Synechocystis* 6803. To corroborate the replication of intact plasmids VIII.23 and VIII.44 in these strains, we re-isolated the plasmids from cyanobacteria and successfully transformed *E. coli* with them (Supplementary Figure S5). Moreover, we detected strong sfGFP expression in the transformants of both cyanobacteria, but not in the respective wild types (Figure 6). The sfGFP accumulation was inducible by the addition of Cu_2SO_4 to a final concentration of 1.25 μM in *Synechocystis* 6803 (Figure 6A), while sfGFP was constitutively present in *Synechococcus* 7942, independent from the absence or presence of added Cu_2SO_4 (Figure 6B). This result is consistent with the presence of a regulatory system that controls the copper-dependent P_{petE} expression in *Synechocystis* 6803, but is lacking in *Synechococcus* 7942 (García-Cañas et al., 2021).

Based on these results, we conclude that the *ssr7036*/asRNA1/*slr7037* locus allows the construction of plasmids that can replicate in

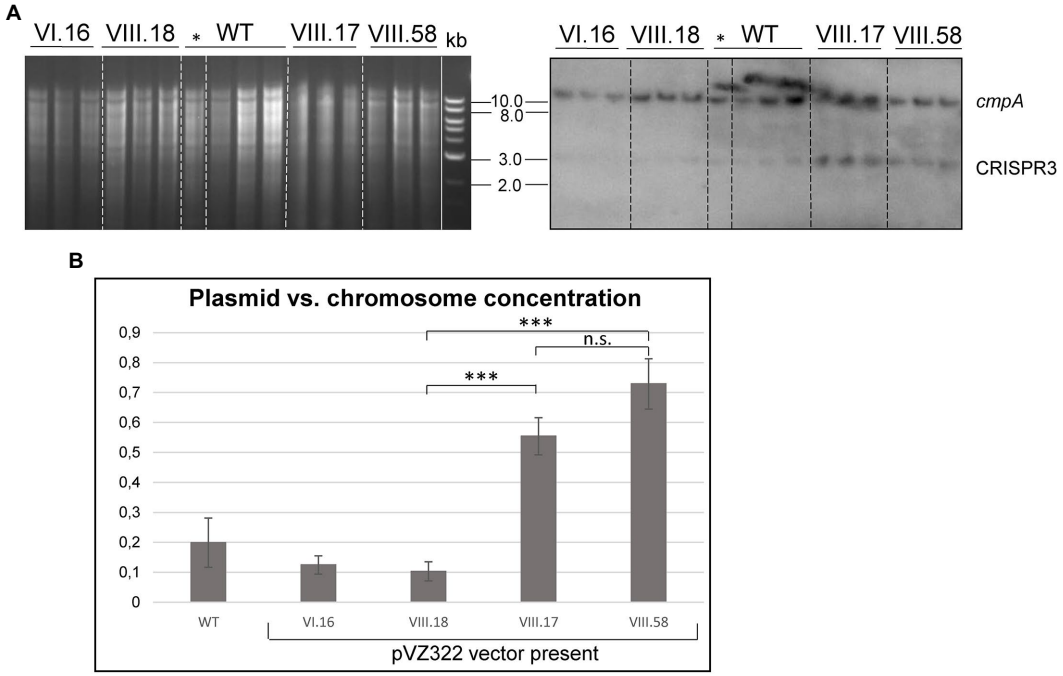

FIGURE 5
Ectopic overexpression of *ssr7036* leads to enhanced levels of pSYSA. **(A)** Southern blot hybridization against *cmpA* and CRISPR3 (right) and the corresponding gel (left). Four µg HindIII-digested DNA were loaded per lane. Expression was induced by 3.3mM rhamnose for 24h except for the one clone of the WT marked with *. WT: wild type DNA; VIII.17: *Synechocystis* 6803+expressing functional *ssr7036* gene; VIII.58: *Synechocystis* 6803+expressing *ssr7036* with premature stop codon; VI.16: *Synechocystis* 6803+pVZ322s control; VIII.18: *Synechocystis* 6803+GFP control. See Table 1 for further details of the used plasmids and strains. Three independent clones were used for each plasmid as biological replicates. An 0.8% agarose gel was used for gel electrophoresis. **(B)** Quantification of the signals. Standard errors from triplicates are shown as error bars. An unpaired two-tailed Student's *t*-test was performed (see Supplementary Table S2 for details). n.s., not significant difference; *** significant difference with value of $p<0.001$.

cyanobacteria and which can be utilized to achieve high expression levels of cargo genes.

Deletion of *slr7037* leads to integration of the entire pSYSA plasmid into the chromosome or into plasmid pSYSX

Based on the observed dependency of *ssr7036*/asRNA1-primed replication on Slr7037, we wondered if *Synechocystis* 6803 could be cured of the native pSYSA plasmid in the absence of *slr7037*. To answer this question, we generated a deletion strain Δ*slr7037* and examined whether pSYSA was retained in it. Three clones with a complete disruption of the *slr7037* gene (Δ*slr7037*-1,2,3) were selected for further analyses (Supplementary Figures S1A,B). Growth of the three clones was comparable to that of the wild type (Supplementary Figure S1C), and sequence analysis showed that they harbored the entire pSYSA sequence (Supplementary Table S1). However, the genomic structure of Δ*slr7037* strains clearly differed from that of the wild type (Figure 7A): Read pairs mapped on pSYSA and the chromosome were detected in Δ*slr7037*-1 and-2, indicating that pSYSA was integrated into the chromosome in these clones (Figure 8). In both Δ*slr7037*-1 and-2, pSYSA was integrated into the chromosome *via* pSYSA genes *slr7104-slr7105*. The fusion sites in the chromosome were *sll0699-sll0700* for Δ*slr7037*-1 and *sll1256-sll1257* for Δ*slr7037*-2. Both *sll0699-sll0700* and *sll1256-sll1257* are homologous genes to *slr7104-slr7105* and encode a transposase. In Δ*slr7037*-2, the number of reads mapped to pSYSA and the chromosome was clearly lower than that observed for the other clones (Figure 8), which seems to be due to a 10-fold increase in the copy number of the small plasmids, pCA2.4, pCB2.4 and pCC5.2 (Supplementary Table S1). There was no evidence for chromosomal integration of pSYSA in the Δ*slr7037*-3 strain. Instead, we observed a fusion between pSYSA and pSYSX (Figure 7) *via* the genes *slr7005* (pSYSA) and *sll6109* (pSYSX). Both genes are identical in sequence and encode putative site-specific integrases. PCR analysis showed complete integration of pSYSA into the chromosome or pSYSX in these strains (Figure 7B). These results indicate that pSYSA was not able to replicate as a separate replicon in the Δ*slr7037* mutants, but that its genes were maintained in the chromosome or pSYSX. Therefore, we suggest that Slr7037 functions as a plasmid-encoded replication initiator protein (Rep) protein and renamed it to CyRepA1 (Cyanobacterial Rep protein A1).

Discussion

A pSYSA-derived plasmid for engineering of cyanobacteria

Here we provide evidence that the 3,566 nt long DNA fragment (positions 32,057 to 35,622 on plasmid pSYSA), which contains the *ssr7036* and *slr7037* genes, enabled a pUC19 derivative to replicate independently in *Synechocystis* 6803 and *Synechococcus* 7942. Thus, this DNA region contains an origin of replication which is functional in these two cyanobacteria. We assume that this might also be the case in further cyanobacterial strains. We show that the resulting vector

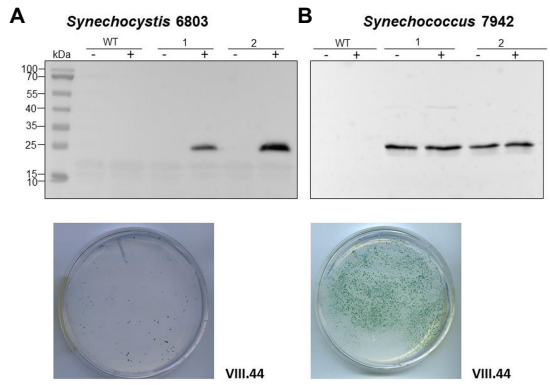

FIGURE 6
An *sfGFP* reporter gene on a pSYSA-derived vector and controlled by the P_{petE} promoter is tightly regulated in *Synechocystis* 6803 and constitutively expressed in *Synechococcus* 7942. **(A)** Expression of sfGFP detected by western blot analysis in *Synechocystis* 6803 after transformation with plasmid VIII.44 which contains the pSYSA origin of replication, i.e., the entire *ssr7036-slr7037* locus. Total protein extracts were prepared from two biological replicates, before and after induction of sfGFP expression by adding Cu_2SO_4 to a final concentration of 1.25µM. The WT was used as negative control. After transformation, selection and re-streaking, the positive clones were grown on plates for 7weeks and for 1week in liquid culture before they were collected for western blot analysis. The plate shows an example 2weeks after transformation with 5µg of VIII.44 DNA (111 colonies). **(B)** Western blot analysis demonstrating the expression of sfGFP in *Synechococcus* 7942. Total protein extracts were prepared from two biological replicates, before and after adding Cu_2SO_4 as in panel **(A)** but the regulatory system for the control of this promoter does not exist in *Synechococcus* 7942 (García-Cañas et al., 2021). After transformation, selection and re-streaking, the positive clones were cultivated on plates for 4weeks and for 2weeks in liquid culture before the analysis. The plate shows a representative example 2weeks after transformation with 1µg of VIII.44 DNA (~1,800 colonies). In both panels **(A,B)**, anti-GFP antiserum (Abcam) was used at a dilution of 1:5,000. PageRuler™ Prestained Protein Ladder was used as marker.

VIII.23 can be engineered to express a reporter gene (plasmid VIII.44; Figure 6). Hence, it can be used as a shuttle vector between cyanobacteria and *E. coli*. The presence of VIII.23 in *Synechocystis* 6803 and in *Synechococcus* 7942 and of VIII.44 in *Synechococcus* 7942 was verified 8 weeks after transformation by re-transformation of plasmid DNA isolated from the cyanobacteria into *E. coli* (plasmid re-isolation, Supplementary Figure S5). Thus, plasmid VIII.23 was maintained despite a possible plasmid incompatibility with the native pSYSA plasmid in *Synechocystis* 6803, at least for this period of time. For *Synechococcus* 7942, these results unequivocally support the applicability of VIII.23-and VIII.44-derived plasmids as shuttle vectors with *E. coli*.

The function of RNase E in the pSYSA copy control mechanism

The initial finding leading to this work was the higher pSYSA copy number observed in strains *rne*(WT) and *rne* (5p) compared to *Synechocystis* WT (Hoffmann et al., 2021). These strains are characterized by an increased level of RNase E relative to the WT, either in its native or 5′ monophosphate sensing deficient form. RNase E is an essential enzyme in cyanobacteria with multiple functions in RNA maturation, degradation and in the post-transcriptional regulation of gene expression (Zhang et al., 2022).

The identification of the *ssr7036*/asRNA1 locus on pSYSA, from which two partially overlapping transcripts originate, provides a link between RNase E and pSYSA copy number control. We show by the analysis of RNA-seq data, *in vitro* cleavage assays and Northern blot hybridizations that both transcripts are processed by RNase E at multiple sites (Figures 2–4). We observed that the most striking effects on the *ssr7036* transcript were caused by cleavage at positions 32,260 and at the twin cleavage sites ~5 nt upstream of the start codon

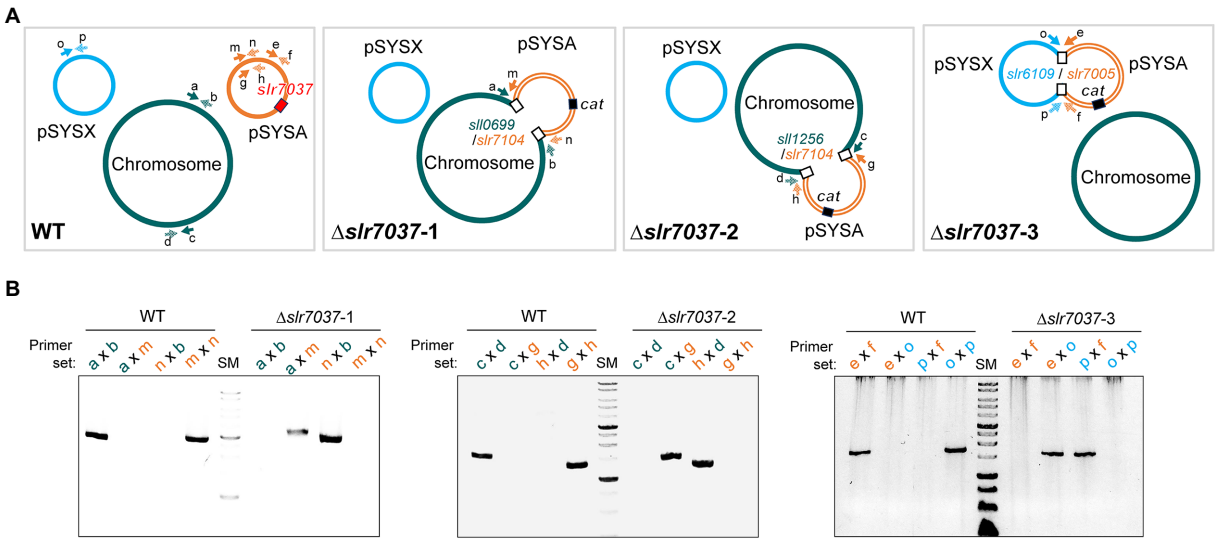

FIGURE 7
Integration of pSYSA into the chromosome or pSYSX through homologous transposase or integrase genes. **(A)** Genome structure of WT and Δ*slr7037* strains as inferred from sequencing analysis. The positions of *slr7037* in WT, the *cat* gene in Δ*slr7037* mutants and fusion positions between pSYSA and chromosome or pSYSX are shown by filled or open boxes, respectively. The pSYSA plasmid is indicated by the orange double lines. **(B)** The fusion sites of pSYSA and chromosome or pSYSX were amplified using genomic DNA extracted from WT and Δ*slr7037* strains as templates with their specific primers shown in panel **A**. The thick bands in the size marker (SM) lanes are 3kb and 1kb, respectively.

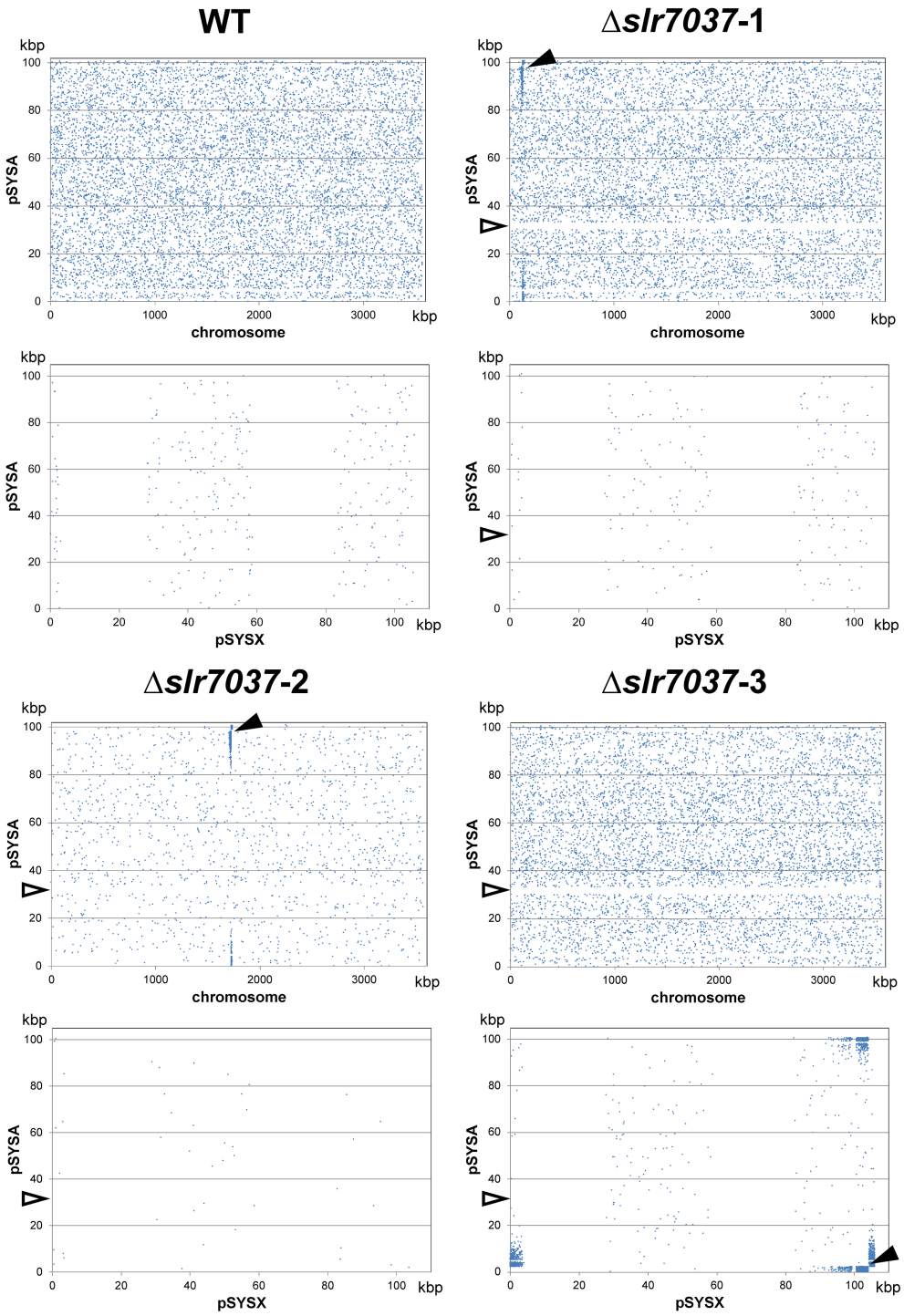

FIGURE 8
The location of paired reads mapped to pSYSA and chromosome (upper panels) or pSYSX (lower panels). Blue dots indicate the position of the start of paired reads mapped to pSYSA and chromosomes or pSYSX. Black arrow heads indicate the fusion points between pSYSA and chromosome or pSYSX. White arrow heads indicate the location of Δ*slr7037*.

and ~ 5 nt into the coding region of *ssr7036* (*cf.* Table 3). These processing events were easily identified in transcriptomic data and these data also showed an inverse effect on the upstream (TU7029) and the downstream (TU7030) segments of the precursor transcript: Whereas TU7029 accumulated to a higher level after the inactivation of the temperature-sensitive RNase E variant at 39°C, TU7030 accumulated to a lower level (Figure 3A). In contrast, the higher amount of RNase E in *rne*(WT) led to the stabilization of TU7030, leading to a higher level of some processing products. The longest of these TU7030 transcripts finish within the *ssr7036-slr7037* intergenic spacer (Figures 1A, 3A,B, upper panels), where pSYSA replication likely is primed.

Comparison of pSYSA replication control with ColE1 and other plasmid systems

Our data suggest that pSYSA is subject to theta plasmid replication, in which the leading strand of a circular plasmid is initiated at a predetermined site (reviewed by (Lilly and Camps, 2015)). In some instances of theta plasmid replication, melting of the DNA double strand depends on transcription, while plasmid-encoded trans-acting Rep proteins can also play a role. Our results show that Slr7037 is essential for maintenance of pSYSA as an autonomous replicon, defining it as a Rep protein. Phylogenetic analysis revealed that homologs of Slr7037 are widely conserved among cyanobacteria (Sakamaki et al., 2022), thus we propose to call this protein CyRepA1. *Synechocystis* 6803 encodes a second Rep protein, ORF B on the small plasmid pCC5.2. pCC5.2 replicates through the rolling-circle mechanism (Xu and McFadden, 1997). Since ORF B does also have replication activity in cyanobacterial cells, we recently proposed to rename it to CyRepA2 (Sakamaki et al., 2022). Intriguingly, even though CyRepA1 and CyRepA2 are involved in different modes of replication, their protein sequences are related (55% similar and 38% identical amino acids) and their predicted 3D structures resemble each other as well (Sakamaki et al., 2022).

The copy number of theta plasmids is controlled at the initiation of replication and frequently involves RNAs in antisense orientation to an RNA that is essential for replication, e.g., by acting as a primer for the initiation of replication. Among the best studied examples for this type of copy number control are the ColE1-like origins of replication, which have been fundamental for the development of various expression vectors. In the *E. coli* ColE1 system, RNAI and RNAII, two overlapping, non-coding RNAs play crucial roles. RNAII is the RNA that primes plasmid DNA replication. The level of RNAII is regulated *via* base-pairing to RNAI, a second, shorter RNA that is much more abundant and sometimes divergently regulated. The TSSs from which RNAI and RNAII originate are 108 bp apart. RNAI has a typical structure consisting of three extended stem-loop elements (Figure 9). The level of RNAI is controlled by RNase E that cleaves a pentanucleotide from its 5′ end, destabilizing it effectively (Morita and Oka, 1979; Lin-Chao and Cohen, 1991). In *Synechocystis* 6803, the TSSs of asRNA1 and *ssr7036* are more distantly localized (304 nt). Judged by the number of reads in prior RNA-seq analyses, asRNA1 is on average more than 100 times more abundant than *ssr7036* (Figure 1). The asRNA1 is almost constitutively present, while the *ssr7036* mRNA level is highly variable and condition-dependent. These findings point at the possible effects of differential transcriptional regulation, which could be addressed in future studies. Compared to RNAI, the asRNA1 is longer (~150 nt compared to 108 nt), but the predicted secondary structure of asRNA1 resembles RNAI by the presence of three stem-loops (Figure 9). Therefore, it is possible that it uses also a similar mechanism to contact its cognate partner molecule, the coding section of the *ssr7036* mRNA, *via* kissing complexes, i.e., base pairing between complementary sequences at the respective loop sections as described for the RNAI:RNAII interaction (Eguchi et al., 1991).

However, there are also major differences between the pSYSA and the ColE1 systems. One of these is the presence of at least two RNase E cleavage sites in asRNA1, which are not close to its 5′ end as in the ColE1 RNAI but within the single-stranded region of the asRNA1 molecule separating stem-loops 1 and 2 from each other (Figure 9).

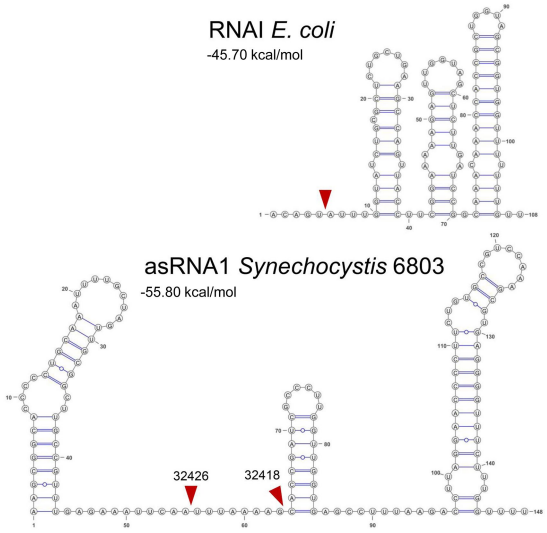

FIGURE 9
Secondary structures and mapped RNase E cleavage sites in the ColE1 RNAI and the pSYSA asRNA1. Both transcripts are with 108 and ~150nt relatively small non-coding RNAs and both overlap a substantial section of another transcript, RNAI in case of the ColE1 replicon and the *ssr7036* mRNA in case of the pSYSA plasmid. Their secondary structures are shown here together with mapped RNase E cleavage sites (read arrowheads) according to this work (asRNA1) or previous analyses for RNAI (Morita and Oka, 1979; Lin-Chao and Cohen, 1991). Sequences of *E. coli* RNAI can be found in the RFAM database (Kalvari et al., 2018) under the accession number RF00106 and the here shown sequence was taken from the European Nucleotide Archive (https://www.ebi.ac.uk/ena), accession no. S42973. The RNA secondary structures were predicted using RNAfold as part of the ViennaRNA Web Services (Gruber et al., 2015) with default parameters, in case of RNAI manually adapted according to published structural models (Morita and Oka, 1979; Lin-Chao and Cohen, 1991) and visualized with VARNA version 3.93 (Darty et al., 2009). The minimum free energy values are indicated in kcal/mol as predicted by RNAfold.

It is likely that these cleavage events destabilize asRNA1. Upon overexpression of RNase E we observed the putative resulting cleavage fragments by Northern blot hybridization. Though the pattern suggested their rapid further turnover (Figure 2B). Another difference is the presence of a small protein, called Rop (repressor of primer), respectively Rom (RNA one modulator), a 63 amino acids protein relevant for ColE1-type replication encoded downstream of the origin of replication (Eguchi and Tomizawa, 1990). Acting as an adaptor protein (Helmer-Citterich et al., 1988), Rom/Rop enhances the binding between RNA I and RNA II, thereby increasing the inhibitory activity of RNA I (Tomizawa and Som, 1984). Accordingly, Rom/Rop defective mutants show an increased plasmid copy number (Vieira and Messing, 1982). Similarly, the pSYSA system includes a small protein (Ssr7036) as well. It consists of 64 amino acids, but unlike Rom/Rop, it is not encoded in the vicinity but within one of the interacting RNAs. Moreover, the predicted secondary structures and the basic IEP of 9.06 differentiates Ssr7036 from the acidic Rom/Rop (Supplementary Figure S6). However, most importantly, we observed that the ectopic expression of a second *ssr7036* gene copy led to a significantly higher pSYSA copy number compared to the chromosome (Figure 5). This effect was also clearly observable if the additional *ssr7036* copy contained a stop codon. Therefore, it was caused by the enhanced transcript level, irrespectively of any hypothetical protein function. The interaction with ribosomes and

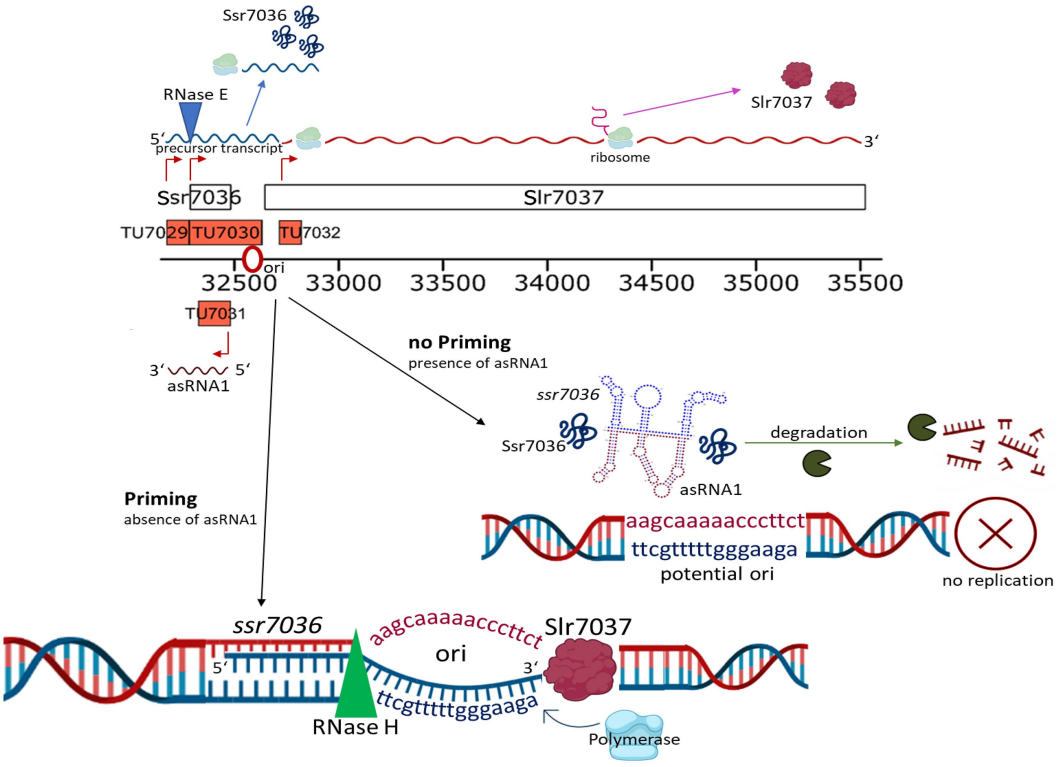

FIGURE 10
The roles of overlapping small RNAs, RNase E and the Slr7037 Rep protein in pSYSA replication. The *ssr7036* mRNA is central in pSYSA replication. It is processed by RNase E into two transcriptional units, TU7029 and TU7030. An abundant transcript, asRNA1, that is complementary to the coding sequence of *ssr7036* accumulates in parallel. Therefore, the protein Ssr7036 can only be produced from mRNA species that are not truncated by RNase E and if translation is not hampered due to interaction with asRNA1. Ssr7036 may be involved in the interaction between the two transcripts but direct evidence is lacking. Those *ssr7036* transcripts that extend beyond the asRNA1 TSS to a position in the *ssr7036-slr7037* intergenic spacer, are likely those that prime pSYSA replication. The downstream located gene *slr7037* encodes the CyRepA1 protein, as defined in this work. The scheme was drawn using elements from the BioRender platform (https://biorender.com/).

translation of *ssr7036* RNA may stabilize and protect the RNA from degradation. Therefore, in contrast to Rom/Rop's inhibitory function on plasmid copy numbers, Ssr7036 seems have a rather additional positive effect on the pSYSA copy number.

Our model summarizes the roles of the two overlapping small RNAs, RNase E and the CyRepA1 protein in pSYSA replication (Figure 10). The *ssr7036* mRNA is transcribed as part of a precursor transcript that is processed by RNase E at two closely spaced major cleavage sites into two transcriptional units, TU7029 and TU7030. Several minor cleavage sites were detected as well. The asRNA1, an abundant transcript that is complementary to the coding sequence of *ssr7036,* accumulates in parallel. Therefore, the protein Ssr7036 can only be produced from mRNA species that are not truncated by RNase E and if translation is not hampered due to interaction with asRNA1. Such *ssr7036* transcripts can extend to a position in the *ssr7036-slr7037* intergenic spacer, where replication likely is primed. The downstream gene *slr7037* encodes a protein with predicted helicase and primase domains, typical for a Rep protein involved in plasmid replication. Indeed, our data suggest that Slr7037 is essential for pSYSA replication. However, the genetic information on pSYSA was not lost in the Δ*slr7037* deletion strains because the entire pSYSA plasmid recombined into the chromosome or the pSYSX plasmid. *Synechocystis* 6803 is highly polyploid (Zerulla et al., 2016). Therefore, the initial strain after the deletion of *slr7037* likely was heterozygous. However, our PCR analysis (Figure 7B) clearly showed that the genotypes in the *cyRepA1* deletion strains (Δ*slr7037*) were homozygous: the plasmid pSYSA was fused to all chromosome or pSYSX copies. This result indicates the presence of a gene conversion mechanism in *Synechocystis* 6803.

That the genetic information present on pSYSA was maintained in the Δ*slr7037* strains is consistent with the presence of at least seven different toxin-antitoxin systems on pSYSA (Kopfmann and Hess, 2013), which should mediate a strong post-segregational killing effect if the genetic information would be lost. Moreover, the results of CRISPRi screening for growth-related genes revealed five genes on pSYSA that are necessary for growth (Yao et al., 2020). These may have contributed to keeping the genetic information in the cell as well.

The gathered information led to the construction of plasmids suitable as shuttle vectors for the genetic manipulation of another model cyanobacterium, *Synechococcus* 7942. These provide a solid basis for the development of further vectors and the manipulation of additional cyanobacterial strains or species.

Data availability statement

The datasets presented in this study can be found in online repositories. The names of the repository/repositories and accession number(s) can be found at: https://ddbj.nig.ac.jp/resource/sra-experiment/DRX398828,

https://ddbj.nig.ac.jp/resource/sra-experiment/DRX398829, https://ddbj.nig.ac.jp/resource/sra-experiment/DRX398830, https://ddbj.nig.ac.jp/resource/sra-experiment/DRX398831.

Author contributions

WH designed the study. AK and VR constructed plasmid vectors, engineered cyanobacterial lines and did molecular analyses. VR performed RNase E *in vitro* assays. CS analyzed transcriptomic data. UH and AW analyzed TIER-seq data. TA, MS, KN-M, and SW constructed *slr7037* deletion strains and analyzed their genomic structure. AK, VR, SW, and WH drafted the manuscript with input from all authors. All authors contributed to the article and approved the submitted version.

Funding

This work was supported by the Deutsche Forschungsgemeinschaft (DFG) Research Training Group MeInBio [322977937/GRK2344] to AW and WH, through the DFG priority program SPP 2002 "Small Proteins in Prokaryotes, an Unexplored World," grant HE2544/12-2 to WH, DFG grant STE 1192/4-2 to CS, and the Ministry of Education, Culture, Sports, Science and Technology of Japan (20K05793) to SW. The mate-pair sequencing was supported by the Cooperative Research Grant of the Genome Research for BioResource, NODAI Genome Research Center, Tokyo University of Agriculture. We acknowledge support by the Open Access Publication Fund of the University of Freiburg.

Acknowledgments

We thank Ingeborg Scholz and Luisa Hemm (both at the University of Freiburg) for providing the RNA samples in Figure 1C and plasmid VII.48, respectively, and John Heap (Imperial College London) for an aliquot of pCK355.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

Supplementary material

The Supplementary material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fmicb.2023.1112307/full#supplementary-material

References

Behler, J., Sharma, K., Reimann, V., Wilde, A., Urlaub, H., and Hess, W. R. (2018). The host-encoded RNase E endonuclease as the crRNA maturation enzyme in a CRISPR–Cas subtype III-Bv system. *Nat. Microbiol.* 3, 367–377. doi: 10.1038/s41564-017-0103-5

Beyer, H. M., Gonschorek, P., Samodelov, S. L., Meier, M., Weber, W., and Zurbriggen, M. D. (2015). AQUA cloning: a versatile and simple enzyme-free cloning approach. *PLoS One* 10:e0137652. doi: 10.1371/journal.pone.0137652

Brantl, S., Birch-Hirschfeld, E., and Behnke, D. (1993). RepR protein expression on plasmid pIP501 is controlled by an antisense RNA-mediated transcription attenuation mechanism. *J. Bacteriol.* 175, 4052–4061. doi: 10.1128/jb.175.13.4052-4061.1993

Darty, K., Denise, A., and Ponty, Y. (2009). VARNA: interactive drawing and editing of the RNA secondary structure. *Bioinformatics* 25, 1974–1975. doi: 10.1093/bioinformatics/btp250

Dühring, U., Axmann, I. M., Hess, W. R., and Wilde, A. (2006). An internal antisense RNA regulates expression of the photosynthesis gene isiA. *Proc. Natl. Acad. Sci. U. S. A.* 103, 7054–7058. doi: 10.1073/pnas.0600927103

Eguchi, Y., Itoh, T., and Tomizawa, J. (1991). Antisense RNA. *Annu. Rev. Biochem.* 60, 631–652. doi: 10.1146/annurev.bi.60.070191.003215

Eguchi, Y., and Tomizawa, J. (1990). Complex formed by complementary RNA stem-loops and its stabilization by a protein: function of ColE1 rom protein. *Cells* 60, 199–209. doi: 10.1016/0092-8674(90)90736-X

García-Cañas, R., Giner-Lamia, J., Florencio, F. J., and López-Maury, L. (2021). A protease-mediated mechanism regulates the cytochrome c6/plastocyanin switch in *Synechocystis* sp. PCC 6803. *Proc. Natl. Acad. Sci. U. S. A.* 118:e2017898118. doi: 10.1073/pnas.2017898118

Gruber, A. R., Bernhart, S. H., and Lorenz, R. (2015). The ViennaRNA web services. *Methods Mol Biol* 1269, 307–326. doi: 10.1007/978-1-4939-2291-8_19

Guimarães, V. A., Le Scornet, A., Khemici, V., Hausmann, S., Armitano, J., Prados, J., et al. (2021). RNase J1 and J2 are host-encoded factors for plasmid replication. *Front. Microbiol.* 12:966. doi: 10.3389/fmicb.2021.586886

Hein, S., Scholz, I., Voß, B., and Hess, W. R. (2013). Adaptation and modification of three CRISPR loci in two closely related cyanobacteria. *RNA Biol.* 10, 852–864. doi: 10.4161/rna.24160

Helmer-Citterich, M., Anceschi, M. M., Banner, D. W., and Cesareni, G. (1988). Control of ColE1 replication: low affinity specific binding of Rop (rom) to RNAI and RNAII. *EMBO J.* 7, 557–566. doi: 10.1002/j.1460-2075.1988.tb02845.x

Hiraga, S., Sugiyama, T., and Itoh, T. (1994). Comparative analysis of the replicon regions of eleven ColE2-related plasmids. *J. Bacteriol.* 176, 7233–7243. doi: 10.1128/jb.176.23.7233-7243.1994

Hoffmann, U. A., Heyl, F., Rogh, S. N., Wallner, T., Backofen, R., Hess, W. R., et al. (2021). Transcriptome-wide in vivo mapping of cleavage sites for the compact cyanobacterial ribonuclease E reveals insights into its function and substrate recognition. *Nucleic Acids Res.* 49, 13075–13091. doi: 10.1093/nar/gkab1161

Hoffmann, U. A., Lichtenberg, E., Rogh, S. N., Bilger, R., Reimann, V., Heyl, F., et al. (2023). The role of the 5′ sensing function of ribonuclease E in cyanobacteria. *BIORXIV* 2023:523895. doi: 10.1101/2023.01.13.523895

Kalvari, I., Nawrocki, E. P., Argasinska, J., Quinones-Olvera, N., Finn, R. D., Bateman, A., et al. (2018). Non-coding RNA analysis using the Rfam database. *Current Protocols Bioinform.* 62:e51. doi: 10.1002/cpbi.51

Kaneko, T., Nakamura, Y., Sasamoto, S., Watanabe, A., Kohara, M., Matsumoto, M., et al. (2003). Structural analysis of four large plasmids harboring in a unicellular cyanobacterium, *Synechocystis* sp. PCC 6803. *DNA Res.* 10, 221–228. doi: 10.1093/dnares/10.5.221

Kaneko, T., Sato, S., Kotani, H., Tanaka, A., Asamizu, E., Nakamura, Y., et al. (1996). Sequence analysis of the genome of the unicellular cyanobacterium *Synechocystis* sp. strain PCC6803. II. Sequence determination of the entire genome and assignment of potential protein-coding regions. *DNA Res.* 3, 109–136. doi: 10.1093/dnares/3.3.109

Kanesaki, Y., Shiwa, Y., Tajima, N., Suzuki, M., Watanabe, S., Sato, N., et al. (2012). Identification of substrain-specific mutations by massively parallel whole-genome resequencing of *Synechocystis* sp. PCC 6803. *DNA Res.* 19, 67–79. doi: 10.1093/dnares/dsr042

Kelly, C. L., Taylor, G. M., Šatkutė, A., Dekker, L., and Heap, J. T. (2019). Transcriptional terminators allow leak-free chromosomal integration of genetic constructs in cyanobacteria. *Microorganisms* 7:263. doi: 10.3390/microorganisms7080263

Kopf, M., Klähn, S., Scholz, I., Matthiessen, J. K. F., Hess, W. R., and Voß, B. (2014). Comparative analysis of the primary transcriptome of *Synechocystis* sp. PCC 6803. *DNA Res.* 21, 527–539. doi: 10.1093/dnares/dsu018

Kopfmann, S., and Hess, W. R. (2013). Toxin-antitoxin systems on the large defense plasmid pSYSA of *Synechocystis* sp. PCC 6803. *J. Biol. Chem.* 288, 7399–7409. doi: 10.1074/jbc.M112.434100

Kopfmann, S., Roesch, S. K., and Hess, W. R. (2016). Type II toxin-antitoxin systems in the unicellular cyanobacterium *Synechocystis* sp. PCC 6803. *Toxins (Basel)* 8:228. doi: 10.3390/toxins8070228

Kuhlemeier, C. J., Thomas, A. A., van der Ende, A., van Leen, R. W., Borrias, W. E., van den Hondel, C. A., et al. (1983). A host-vector system for gene cloning in the cyanobacterium *Anacystis nidulans* R2. *Plasmid* 10, 156–163. doi: 10.1016/0147-619x(83)90068-9

Lilly, J., and Camps, M. (2015). Mechanisms of theta plasmid replication. *Microbiol. Spec.* 3:3.1.02. doi: 10.1128/microbiolspec.PLAS-0029-2014

Lin-Chao, S., and Cohen, S. N. (1991). The rate of processing and degradation of antisense RNAI regulates the replication of ColE1-type plasmids in vivo. *Cells* 65, 1233–1242. doi: 10.1016/0092-8674(91)90018-t

Love, M. I., Huber, W., and Anders, S. (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* 15:550. doi: 10.1186/s13059-014-0550-8

Maeda, K., Okuda, Y., Enomoto, G., Watanabe, S., and Ikeuchi, M. (2021). Biosynthesis of a sulfated exopolysaccharide, synechan, and bloom formation in the model cyanobacterium *Synechocystis* sp. *Strain PCC* 10:e66538. doi: 10.7554/eLife.66538

Malmgren, C., Engdahl, H. M., Romby, P., and Wagner, E. G. (1996). An antisense/target RNA duplex or a strong intramolecular RNA structure 5′ of a translation initiation signal blocks ribosome binding: the case of plasmid R1. *RNA* 2, 1022–1032. PMID: 8849778

Malmgren, C., Wagner, E. G., Ehresmann, C., Ehresmann, B., and Romby, P. (1997). Antisense RNA control of plasmid R1 replication. The dominant product of the antisense RNA-mRNA binding is not a full RNA duplex. *J. Biol. Chem.* 272, 12508–12512. doi: 10.1074/jbc.272.19.12508

Mitschke, J., Georg, J., Scholz, I., Sharma, C. M., Dienst, D., Bantscheff, J., et al. (2011). An experimentally anchored map of transcriptional start sites in the model cyanobacterium *Synechocystis* sp. PCC6803. *Proc. Natl. Acad. Sci. U. S. A.* 108, 2124–2129. doi: 10.1073/pnas.1015154108

Morita, M., and Oka, A. (1979). The structure of a transcriptional unit on colicin E1 plasmid. *Eur. J. Biochem.* 97, 435–443. doi: 10.1111/j.1432-1033.1979.tb13131.x

Pinto, F., Thapper, A., Sontheim, W., and Lindblad, P. (2009). Analysis of current and alternative phenol based RNA extraction methodologies for cyanobacteria. *BMC Mol. Biol.* 10:79. doi: 10.1186/1471-2199-10-79

Reimann, V., Alkhnbashi, O. S., Saunders, S. J., Scholz, I., Hein, S., Backofen, R., et al. (2017). Structural constraints and enzymatic promiscuity in the Cas6-dependent generation of crRNAs. *Nucleic Acids Res.* 45, 915–925. doi: 10.1093/nar/gkw786

Rippka, R., Deruelles, J., Waterbury, J. B., Herdman, M., and Stanier, R. Y. (1979). Generic assignments, strain histories and properties of pure cultures of cyanobacteria. *Microbiology* 111, 1–61. doi: 10.1099/00221287-111-1-1

Sakamaki, Y., Maeda, K., Nimura-Matsune, K., Chibazakura, T., and Watanabe, S. (2022). Exploration of the autonomous replication region and its utilization for expression vectors in cyanobacteria. *BIORXIV* 2022:516977. doi: 10.1101/2022.11.17.516977

Scholz, I., Lange, S. J., Hein, S., Hess, W. R., and Backofen, R. (2013). CRISPR-Cas systems in the cyanobacterium *Synechocystis* sp. PCC6803 exhibit distinct processing pathways involving at least two Cas6 and a Cmr2 protein. *PLoS One* 8:e56470. doi: 10.1371/journal.pone.0056470

Tomizawa, J., and Som, T. (1984). Control of cole 1 plasmid replication: enhancement of binding of RNA I to the primer transcript by the rom protein. *Cells* 38, 871–878. doi: 10.1016/0092-8674(84)90282-4

Trautmann, D., Voß, B., Wilde, A., Al-Babili, S., and Hess, W. R. (2012). Microevolution in cyanobacteria: re-sequencing a motile substrain of *Synechocystis* sp. PCC 6803. *DNA Res.* 19, 435–448. doi: 10.1093/dnares/dss024

Vieira, J., and Messing, J. (1982). The pUC plasmids, an M13mp7-derived system for insertion mutagenesis and sequencing with synthetic universal primers. *Gene* 19, 259–268. doi: 10.1016/0378-1119(82)90015-4

Xu, W., and McFadden, B. A. (1997). Sequence analysis of plasmid pCC5.2 from cyanobacterium *Synechocystis* PCC 6803 that replicates by a rolling circle mechanism. *Plasmid* 37, 95–104. doi: 10.1006/plas.1997.1281

Yang, X., and McFadden, B. A. (1993). A small plasmid, pCA2.4, from the cyanobacterium *Synechocystis* sp. strain PCC 6803 encodes a rep protein and replicates by a rolling circle mechanism. *J. Bacteriol.* 175, 3981–3991. doi: 10.1128/jb.175.13.3981-3991.1993

Yang, X., and McFadden, B. A. (1994). The complete DNA sequence and replication analysis of the plasmid pCB2.4 from the cyanobacterium *Synechocystis* PCC 6803. *Plasmid* 31, 131–137. doi: 10.1006/plas.1994.1014

Yanisch-Perron, C., Vieira, J., and Messing, J. (1985). Improved M13 phage cloning vectors and host strains: nucleotide sequences of the M13mp18 and pUC19 vectors. *Gene* 33, 103–119. doi: 10.1016/0378-1119(85)90120-9

Yao, L., Shabestary, K., Björk, S. M., Asplund-Samuelsson, J., Joensson, H. N., Jahn, M., et al. (2020). Pooled CRISPRi screening of the cyanobacterium *Synechocystis* sp PCC 6803 for enhanced industrial phenotypes. *Nat. Commun.* 11:1666. doi: 10.1038/s41467-020-15491-7

Zerulla, K., Ludt, K., and Soppa, J. (2016). The ploidy level of *Synechocystis* sp. PCC 6803 is highly variable and is influenced by growth phase and by chemical and physical external parameters. *Microbiology* 162, 730–739. doi: 10.1099/mic.0.000264

Zhang, J.-Y., Hess, W. R., and Zhang, C.-C. (2022). "Life is short, and art is long": RNA degradation in cyanobacteria and model bacteria. *mLife* 1, 21–39. doi: 10.1002/mlf2.12015

Zhang, L., McSpadden, B., Pakrasi, H. B., and Whitmarsh, J. (1992). Copper-mediated regulation of cytochrome c553 and plastocyanin in the cyanobacterium *Synechocystis* 6803. *J. Biol. Chem.* 267, 19054–19059. doi: 10.1016/S0021-9258(18)41739-5

Zinchenko, V. V., Piven, I. V., Melnik, V. A., and Shestakov, S. V. (1999). Vectors for the complementation analysis of cyanobacterial mutants. *Russ. J. Genet.* 35, 228–232.

frontiers | Frontiers in Microbiology

TYPE Mini Review
PUBLISHED 02 March 2023
DOI 10.3389/fmicb.2023.1143770

OPEN ACCESS

EDITED BY
Jin Liu,
Peking University,
China

REVIEWED BY
Zixi Chen,
Shenzhen University,
China

*CORRESPONDENCE
Pachara Sattayawat
✉ pachara.sattayawat@cmu.ac.th

SPECIALTY SECTION
This article was submitted to Microbiotechnology, a section of the journal Frontiers in Microbiology

RECEIVED 13 January 2023
ACCEPTED 06 February 2023
PUBLISHED 02 March 2023

CITATION
Inwongwan S, Pekkoh J, Pumas C and Sattayawat P (2023) Metabolic network reconstruction of *Euglena gracilis*: Current state, challenges, and applications. *Front. Microbiol.* 14:1143770. doi: 10.3389/fmicb.2023.1143770

Metabolic network reconstruction of *Euglena gracilis*: Current state, challenges, and applications

Sahutchai Inwongwan[1,2], Jeeraporn Pekkoh[1,2], Chayakorn Pumas[1,3] and Pachara Sattayawat[1,2,3]*

[1]Department of Biology, Faculty of Science, Chiang Mai University, Chiang Mai, Thailand, [2]Research Center of Microbial Diversity and Sustainable Utilizations, Faculty of Science, Chiang Mai University, Chiang Mai, Thailand, [3]Research Center in Bioresources for Agriculture, Industry and Medicine, Chiang Mai University, Chiang Mai, Thailand

A metabolic model, representing all biochemical reactions in a cell, is a prerequisite for several approaches in systems biology used to explore the metabolic phenotype of an organism. Despite the use of *Euglena* in diverse industrial applications and as a biological model, there is limited understanding of its metabolic network capacity. The unavailability of the completed genome data and the highly complex evolution of *Euglena* are significant obstacles to the reconstruction and analysis of its genome-scale metabolic model. In this mini-review, we discuss the current state and challenges of metabolic network reconstruction in *Euglena gracilis*. We have collated and present the available relevant data for the metabolic network reconstruction of *E. gracilis*, which could be used to improve the quality of the metabolic model of *E. gracilis*. Furthermore, we deliver the potential applications of the model in metabolic engineering. Altogether, it is supposed that this mini-review would facilitate the investigation of metabolic networks in *Euglena* and further lay out a direction for model-assisted metabolic engineering.

KEYWORDS

metabolic network reconstruction, genome-scale metabolic model, metabolic network modelling, metabolic flux analysis, *Euglena gracilis*

1. Introduction

Euglena gracilis is a photosynthetic protist with a long history of being a model organism in biological studies. It can grow in autotrophic, heterotrophic or mixotrophic conditions (Takeyama et al., 1997), aerobically or anaerobically, and over a wide range of pH (Yamane et al., 2001). *E. gracilis* has been considered as a potential dietary supplement due to its capacity to produce various bioactive compounds and is a useful source of proteins, polyunsaturated fatty acids, vitamin A, vitamin C and vitamin E (Korn, 1964; Takeyama et al., 1997; Kusmic et al., 1998; Ogbonna et al., 1998; Barsanti et al., 2000; Gissibl et al., 2019). Moreover, *E. gracilis* accumulates storage carbohydrate in the form of β-1,3-glucan, which can make up to 85% of cell dry weight (Inui et al., 1982), called paramylon. Several medicinal properties of paramylon have been reported including antiviral and immunomodulatory effects (El Khoury et al., 2012; Murphy et al., 2020). Even though, this organism has long been a focus of research for its diverse industrial applications, it is confounding that the understanding of its metabolic capacity is still highly limited. Moreover, the analysis of the *Euglena* genome is still incomplete and restricted

by its size and complexity (Ebenezer et al., 2019) arisen from multiple secondary endosymbiosis events during its evolution (Novák Vanclová et al., 2020).

E. gracilis has received significant attention as a potential cell factory due to their ability to produce a diverse array of valuable chemicals. Efforts are currently focused on improving their ability to produce these chemicals in a cost-effective manner. Metabolic network reconstruction is an approach used to identify and characterise metabolic pathways present inside of an organism, which allow understandings of the metabolism inside of the cell. These reconstructed metabolic networks can then be used in several aspects. For example, from a metabolic engineering perspective, the comprehensive understanding of the metabolic pathways within a cell enables the rational selection of engineering targets. Interestingly, this approach has been well-explored in other model organisms, yet the development has been slow in the case of *E. gracilis*. Therefore, in this mini-review, the current state of the reconstruction of the metabolic network of *E. gracilis* is established, providing an overview of its metabolic network and highlighting the unique features of the network. In addition, we emphasise the challenges for reconstructing the network model of *E. gracilis* and deliver the available data that could be exploited to improve the completion of the metabolic model. Furthermore, possible applications of the model in metabolic engineering for the production of valuable products are also discussed. An overview of the content is presented in Figure 1.

2. Current state of the reconstruction of the metabolic network of *Euglena gracilis*

The exceptionally versatile metabolic capacity of *E. gracilis* is reflected in its broad range of growth conditions and substrate utilisation. The central pathways have been characterised, including glycolysis, gluconeogenesis, the tricarboxylic acid cycle (TCA), the pentose phosphate pathway (PPP), and the metabolism of lipids and amino acids (Inwongwan et al., 2019). The genome of *Euglena* was estimated to be approximately 500 Mb in size (Ebenezer et al., 2019) and has not been completely analysed. A complete sequence of *E. gracilis* chloroplast genome was published in 1993 (Hallick et al., 1993). The transcriptomic analysis indicates the presence of the biosynthesis pathways of carotenoids, thylakoid glycolipids, fatty acids, and isoprenoids. It also demonstrates the capacity to utilise the pathways for vitamin C, vitamin E, and glutathione metabolism to respond to stresses and to produce multifunctional polydomain proteins related to fungi and bacteria (O'Neill et al., 2015b). A study of the regulatory system of wax-ester metabolism under anaerobic conditions of *E. gracilis* using the comparative transcriptomic approach (Yoshida et al., 2016) reported that the differentially expressed genes from aerobic and anaerobic conditions were not involved in wax-ester metabolism, indicating that the metabolic pathways involved in wax-ester biosynthesis were regulated at the post-transcriptional level. In addition, the following published transcriptome

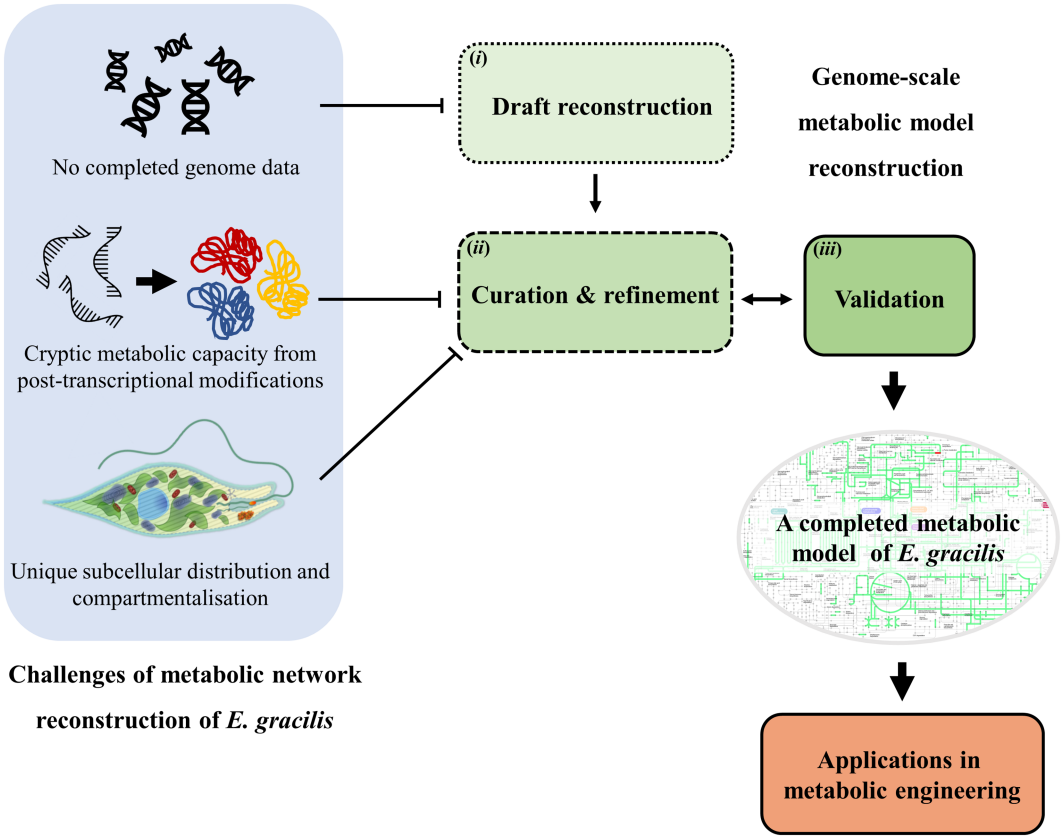

FIGURE 1
A schematic diagram representing the overview of this mini-review including steps of metabolic model reconstruction using genome data: (i) Draft reconstruction (ii) curation and refinement (iii) Validation of the model and challenges encountered when reconstructing the genome-scale model of *Euglena gracilis*. Potential applications of the network model are also discussed.

data of *E. gracilis* also suggested that gene regulation in euglenozoans is not primarily controlled at the transcriptional level (Cordoba et al., 2021). The study of the mitochondrial genome of *E. gracilis* revealed the flexible mitochondrial metabolisms (Dobáková et al., 2015; Ebenezer et al., 2019), the mitochondria can produce energy under either aerobic or anaerobic conditions, and efficiently utilise a diverse set of organic respiratory substrates facilitated by the unique subcellular localisation of the metabolic pathways in mitochondria, such as glyoxylate cycle and alcohol oxidisation (Inwongwan et al., 2019). Altogether, the transcriptomic analyses of *E. gracilis* emphasise its versatile metabolic capacity and the regulation at post-transcriptional level (Yoshida et al., 2016; O'Neill et al., 2015b). Thus, using only the transcriptomic approach might not be sufficient to understand how *E. gracilis* responds to various conditions. The *E. gracilis* plastid proteome indicates the function of photosynthesis and demonstrates the core plastid metabolic pathways (Ebenezer et al., 2019; Novák Vanclová et al., 2020); however, there is no evidence of the presence of oxidative pentose phosphate pathway in its secondary chloroplast. Moreover, metabolomic analysis of *E. gracilis* reported changes in pathways used in response to environmental stresses (He et al., 2021). Nevertheless, the metabolome is not able to elucidate the systemic metabolic operation of the whole metabolic network.

Transcript sequences and topology gap filling were used to attempt to reconstruct the metabolic network of *E. mutabilis* (Halter et al., 2015; Prigent et al., 2017). This network model was reported to be incomplete as it could not simulate the growth of *E. mutabilis* in the dark (Prigent et al., 2017), despite the capacity of this species to do so. A draft of metabolic network model of *E. gracilis* was constructed specifically to study the heterotrophic metabolism of various kinds of carbon substrates which mainly includes the operation of the central metabolic pathways (Inwongwan, 2021), most of the peripheral pathways and reactions of photosynthesis were not extensively curated or tested. There have not been any other reports of a completed reconstructed metabolic network of *E. gracilis*.

3. Challenges of metabolic network reconstruction of *Euglena gracilis*

Genome sequence is generally a prerequisite for reconstructing the metabolic network of an organism (Thiele and Palsson, 2010). Genome-scale metabolic model (GEM) is a mathematical model consists of all metabolic reactions in a cell and their stoichiometries generally based on genome data, which is able to quantify the genotype–phenotype relationships (Fang et al., 2020). It has become a powerful tool in systems biology to study responsive metabolic phenotypes and optimise the production of targeted metabolites in metabolic engineering. Steps of reconstructing GEM start with drafting a reconstruction from the annotated genome data, and then, the draft model is manually curated and refined based on the physiological and/or biochemical evidences to increase the precision and accuracy. Subsequently, the experimental data including biomass composition, media composition and consumption rate, growth characteristics and other environmental factors in the mathematical model format are integrated into the model. After completing the draft reconstruction, network verification, debugging and gap filling are performed. The last step is evaluation and validation of the model depending on the objective of the reconstruction (Thiele and Palsson, 2010). This bottom-up reconstruction procedure applies well with the model organisms due to the accessible organism-specific genome and biochemical data. Several automated tools for generating GEM were developed based on the available databases of the model species, and GEM modelling are particularly well-developed in prokaryotes. However, the metabolic network reconstruction processes are not as straightforward for eukaryotic and non-model species (Yan and Fong, 2017; Hanna et al., 2020), especially for ones with complex evolution causing the diverse and cryptic metabolisms, unique subcellular localisation of pathways and organelles, and without an available completed set of genome data, like *Euglena*. Allegedly, reconstructing the GEM of *E. gracilis* comes with several challenges that need to be overcome to enable the generation of a descriptive GEM. Here, we have listed the main challenges encountered over the years.

3.1. No completed genome data available

Genomic analysis can lead to an investigation of the organism in numerous aspects (Griffiths et al., 2015). The conventional GEM reconstruction requires the completed analysed genome data as the initiating material for drafting the reconstruction as annotated genomes provides the absolute genetic and metabolic capacity of the network. Even though there are several computational automate tools to generate GEMs from genome data, the ability to produce the high-quality GEM and the application of these tools are usually restricted to the well-define organisms, partially due to a lack of complete annotated genome sequence and available related data (Passi et al., 2021). Despite the history of *E. gracilis* in biochemical and physiological research, the genome of *E. gracilis* has not yet to be fully analysed. Its complex evolution results in a massive genome size with a chimeric and convoluted structure, obstructing the assembly and analysis of the genome (Ebenezer et al., 2019; O'Neill et al., 2015a).

3.2. The cryptic metabolic capacity from post-transcriptional modifications

Several transcriptomic analyses underline that *E. gracilis* metabolic phenotypes can be significantly controlled by the post-transcription modification and regulation processes (O'Neill et al., 2015b; Yoshida et al., 2016; Ebenezer et al., 2019; Cordoba et al., 2021), demonstrating the cryptic but great metabolic capacity and complex cellular regulatory mechanisms. The high level of involvement of post-transcriptional modification creates a great challenge in curating and simulating the GEM for the specific conditions of interest. Regardless of the unavailability of the complete genome data, the transcriptomes would be insufficient to indicate the responsive metabolic mechanisms to the condition of interest or to depict the wholistic metabolic capacity of *E. gracilis*. Without this information, some significant metabolic processes in the network could be missed to identify.

3.3. Unique subcellular distribution and compartmentalisation of the metabolic pathways

Reconstruction of GEM of eukaryotes are challenging by the size of genomes and the multitude of cellular compartments (Thiele and

TABLE 1 Published transcriptome, organelle genome and proteome data of *E. gracilis*.

Analysis	Growth condition	Number of component	Reference
Transcriptome	High nutrient: EG + JM media with 15 gL^{-1} glucose, 30°C, 200 rpm, in the dark Low nutrient: $CaCl_2$ (0.1 $g\,L^{-1}$), $NaOAc{\cdot}(H_2O)_3$ (1 $g\,L^{-1}$) and JM medium with 15 gL^{-1} agar, 21°C, ambient light	32,128 unique protein-encoding genes	O'Neill et al. (2015b)
	Koren-Hutner medium, 26°C, 120 rpm, 100 µmol photons m^{-2} s^{-1}, stationary phase cells were incubated in anaerobic condition for 5 min.	26,479 unique protein-encoding genes	Yoshida et al. (2016)
	Dark grown: Hutner medium, ambient temperature, in the dark Light grown: Hutner medium, ambient temperature, illumination from a 60-W tungsten filament bulb at 20 cm from the culture vessel	36,526 unique protein-encoding genes	Ebenezer et al. (2019)
	Liquid mineral medium tris-minimum-phosphate with vitamin mixture supplemented, pH 7.0, 25°C (i) acetate (60 mM) added, in the dark (ii) acetate (60 mM) added, low PPFD (50 µE m^{-2} s^{-1}) (iii) acetate (60 mM) added, medium PPFD (200 µE m^{-2} s^{-1}) (iv) No acetate added, low PPFD (50 µE m^{-2} s^{-1})	49,922 unique protein-encoding genes	Cordoba et al. (2021)
	EG medium, pH 3.5, 21°C, 96 µmol photons m^{-2} s^{-1} for a photoperiod of 16 h light/8 h dark, control and 5 µmol L^{-1} $Hg(NO_3)_2$ treated	439,129 assembled genes	Mangal et al. (2022)
Chloroplast genome	Standard cultivation procedures, unspecified	55 annotated genes	Hallick et al. (1993)
Mitochondrial genome	Hutner medium, 27°C, constant shaking, permanent light conditions 10 µm/m^{-2} s^{-1}	7 annotated genes	Dobáková et al. (2015)
Proteome	Dark grown: Hutner medium, ambient temperature, in the dark Light grown: Hutner medium, ambient temperature, illumination from a 60-W tungsten filament bulb at 20 cm from the culture vessel	8,661 proteins	Ebenezer et al. (2019)
	GNY medium, 23°C, 150 rpm, white light (2000 lx) for a photoperiod of 12 h light/12 h dark, heavy metals treated: mercury (as $HgCl_2$), lead (as $Pb(NO_3)_2$) and cadmium (as $CdCl_2$).	5,325 proteins	Khatiwada et al. (2020)
	Wild-type and Oflaxocin bleached strains, EM medium with 1% ethanol, 25°C, 50 µmol photons m^{-2} s^{-1}	1,572 proteins	Chen et al. (2022)
Plastid proteome	Unspecified	1,345 proteins	Novák Vanclová et al. (2020)
Mitochondrial proteome	Non-phtosyntheisc mutant (strain SM-ZK), Koren–Hutner (KH) medium, 26°C, 120 rpm, under continuous light conditions of 50 µmol photons m^{-2} s^{-1}	714 proteins	Tamaki et al. (2020)
	Hutner medium, 27°C, constant shaking, permanent light conditions 10 µm/m^{-2} s^{-1}	2,704 proteins	Hammond et al. (2020)

Palsson, 2010). The compartmentation of metabolism in eukaryotes, especially ones with plastid (s), complicates the structure on the GEM due to the uncertainties of the distribution of the specific enzymes (Kruger and Ratcliffe, 2015). Locating subcellular locations of proteins and integrating them into the GEMs of eukaryotes have been one of the crucial steps to generate the accurate GEM. The subcellular localisation of central metabolic pathways of *E. gracilis* was previously reported, demonstrating the specific subcellular pathway distribution and the ambiguity for the function of its secondary chloroplast in heterotrophic metabolism. The study also emphasises the difficulty in predicting the subcellular location of *E. gracilis* enzymes from peptide sequences as the transportation into *E. gracilis* chloroplasts is not fully understood or well-characterised (Inwongwan et al., 2019). To complete the GEM of *E. gracilis*, identifying the subcellular locations and functions of all metabolic pathways would greatly flavour the improvement of the network.

As the limited availability of the data is one of the main challenges, all reported data is collated and present in this section (Table 1). As mentioned, the genome of *E. gracilis* has not been completely analysed; thus, reconstructing the GEM of *E. gracilis* is merely possible. A metabolic network model of *E. gracilis* was generated based on high quality transcriptomic data (O'Neill et al., 2015b; Yoshida et al., 2016) and is used in the same sense as GEM. The model was able to predict the metabolic fluxes of the central metabolic pathways during the heterotrophic metabolism of various carbon substrates, but the model has not been further developed or validated to predict the metabolic phenotypes in any other growth conditions (Inwongwan, 2021). However, this transcriptomic-based metabolic network model of *E. gracilis* shows a possibility to reconstruct a functioning metabolic network model without the complete genome data.

The metabolic model reconstruction is an iterative process that should be continuously adjusted with the newly available data to improve the accuracy and completeness of the model. In the light of high throughput analysis of multiomics data, the extensive availability of transcriptome and proteome data could increase the reliability of the GEM of *E. gracilis*. Several approaches were developed to integrate the omics data with GEM. For example, based on seeking steady states of regulatory network, FlexFlux combine the analysis of regulatory

networks with flux balance analysis (FBA) of GEM (Marmiesse et al., 2015), and Metabolic and Expression models (ME-models) includes metabolic and transcriptomic expression with the analysis of GEM (Lerman et al., 2012). Besides, the available transcriptome and proteome data of *E. gracilis* could be used to improve the draft reconstruction of the GEM. The transcriptomes of *E. gracilis* from various conditions provide potential metabolic capacity that could be used to draft a comprehensive GEM. The proteomes of *E. gracilis* provide further insights of the metabolic operation specific to the conditions of interest. These data could help refine the GEM to overcome the cryptic post-transcriptional regulation process. In addition, to constraint or validate the accuracy of the model, numerous extensive biochemical and physiological data of *E. gracilis* are required, such as biomass composition, growth characteristics, non-growth associated maintenance, carbon conversion efficiency, overall metabolic rate, metabolome and mitochondrial physiology. Some of these data are seldomly studied and reported, thus, the currently available data potentially contributing to the reconstruction of a complete GEM of *E. gracilis* might still be far from sufficient and will need to be further analysed.

4. Potential applications of the *Euglena* model in metabolic engineering

Euglena, as mentioned, is used to synthesise a number of high-value compounds (Gissibl et al., 2019). Though several bioproducts are commercially available, strain improvement to allow cost-competitive production is still important. Low production yield is one of the challenges that slows down the commercialisation of these natural products and this could be because the metabolic flux or the flow of metabolites to the desirable final products is low (Liu et al., 2017). Metabolic engineering is, therefore, of interest, as this approach could increase the flux towards desirable products in a stepwise manner by manipulating the expression of bottleneck genes. In order to do so, information regarding *Euglena* metabolic networks and bottleneck reactions towards the target products is essential to ensure successful engineering – which could be implemented by knocking out, overexpressing or heterologously expressing of particular genes. As the development of genetic engineering in *Euglena* has just been kicked off, not a lot of works have been published, and only a couple of reviews summarising the reported engineering tools are available (Harada et al., 2020; Khatiwada et al., 2020; Chen et al., 2022). The delayed development could be due to the lack of knowledge on their molecular characteristics, including their complete nuclear genome sequences. Moreover, they have distinct characteristics such as their chloroplasts that are surrounded with three enveloping membranes, which makes it challenging for DNA transformation. Cellular characteristics important for engineering are often addressed along with the development of compatible engineering tools. Antibiotic resistance is one of the important properties addressed when developing transformation techniques to identify suitable selectable markers (Khatiwada et al., 2019). A chloroplast transformation technique was developed in 2001 using biolistic transformation with confirmed transgene transcription (Doetsch et al., 2001). Later on, the focus was shifted to the nucleus, electroporation was developed and demonstrated to be a potential technique to transform fluorescent markers into the nucleus of *E. gracilis* (Ohmachi et al., 2016). Recently, a nuclear transformation technique with the help of *Agrobacterium* was also demonstrated to be successful in *E. gracilis* (Khatiwada et al., 2019; Becker et al., 2021). RNA interference (RNAi), a technique to suppress gene expression, was also investigated in *Euglena*. As a metabolic engineering strategy, a few reports have used RNAi to silence genes encoding enzymes in competitive pathways for natural product production in *Euglena* (Nakazawa et al., 2015; Kato et al., 2017; Kimura and Ishikawa, 2018). Even though the development of genetic engineering tools in *Euglena* has been relatively slow compared to other model organisms, the number of published works has gradually increased, including the first report of a groundbreaking tool, CRISPR, in 2019 (Nomura et al., 2020). This indicates increased attention towards *Euglena* as a potential host for genetic engineering and bioproduction, as the most recent development of CRISPR system in *Euglena* was to create a non-motile mutant to facilitate the harvesting process (Ishikawa et al., 2022).

As the number of genetic engineering toolkits for *Euglena* has been increasing over time, this ensures the feasibility to metabolically engineer them as a cell factory for attractive chemicals and with the help of metabolic network models, metabolic bottlenecks can be identified. To the best of our knowledge, model-assisted metabolic engineering in *Euglena* has yet to be reported. However, examples have been successfully demonstrated in other hosts, including *Escherichia coli* and *Saccharomyces cerevisiae*. To provide some examples, in 2018, a metabolic model for hydrocarbon production in *E. coli* was reconstructed, and flux balance analysis (FBA), was used to identify metabolic engineering strategies to increase the production of long-chain alkanes and alcohols (Fatma et al., 2018). Recently, a report also utilised FBA to force the flux towards *n*-butane in *E. coli*, which was found to vastly increase the production by 168 folds (Liu et al., 2022). Similarly, in the case of *S. cerevisiae*, FBA was used to identify the target to fine-tune central carbon metabolism to increase the levels of acetyl-CoA and malonyl-CoA (Ferreira et al., 2019). To provide more relatable examples, GEM has also been constructed in photosynthetic organisms. Several metabolic models of cyanobacteria (i.e., *Synechocystis* sp. PCC 6803, *Synechococcus* sp. PCC 7002 and *Arthrospira platensis*) have been reconstructed and summarised in a previous review (Santos-Merino et al., 2019). Several works have reported on the use model-assisted metabolic engineering to improve the production of bioproducts from cyanobacteria including limonene (Wang et al., 2016), 1,3-propanediol (Hirokawa et al., 2017), ethanol (Yoshikawa et al., 2017) and *n*-butanol (Anfelt et al., 2015). Compared to prokaryotic cells, the field of metabolic modeling in eukaryotic photosynthetic organisms has progressed slowly due to the complexity of their massive genome size and cellular compartmentation (Kruger and Ratcliffe, 2015). However, models of Arabidopsis and tomato, two model organisms, have been successfully constructed (Poolman et al., 2009; Yuan et al., 2016) and proven useful for predicting the metabolic phenotype of the organisms. In the case of eukaryotic microalgae, metabolic models for several microalgae, including *Chlorella* and *Chlamydomonas*, have been reconstructed (Chang et al., 2011; Kliphuis et al., 2012; Zuñiga et al., 2016; Tibocha-Bonilla et al., 2018). These models have mostly been used in order to understand the native metabolism of the microalgae. A recent study has reported the use of the metabolic model of *Chlorella vulgaris* to predict cultivating conditions for growth optimisation. Interestingly, the predicted conditions also led to increased production of fatty acid methyl ester

(FAME) and lutein (Li et al., 2019), demonstrating the potential of using metabolic models to enhance bioproduction in microalgae. Redirection of flux from β-1,3-glucan biosynthetic pathway to other pathways could be an approach that would allow increased production of wide range chemicals in *Euglena*. To be specific, according to the flux maps of heterotrophic metabolism from ^{13}C metabolic flux analysis, *Euglena* tends to direct glucose intake (37–41%) towards β-1,3-glucan storage (Inwongwan, 2021). To optimise the production of lipid production with this information, for example, down regulation of paramylon synthetase, an enzyme responsible for β-1,3-glucan synthesis from UDP-glucose, could be the potential strategy. Altogether, from a metabolic engineering point of view, it could be concluded that metabolic network models are valuable for rational design engineering.

Conclusion

This mini-review has summarised the current state of metabolic network reconstruction in *E. gracilis* and the challenges that obstruct the progression of the model. Generating a definitive metabolic model of *E. gracilis* could significantly contribute to the application of this organism as a cell factory for production of valuable compounds. The model can be used to study the metabolism of *E. gracilis* in various conditions and to predict targets for metabolic engineering. Its potential has been demonstrated in other model organisms, yet model-assisted engineering has never been reported in *Euglena*. This could be because, apart from the unavailability completed model of *Euglena*, the delayed development of genetic engineering toolkits though rapid development has been observed over recent years. Moreover, in this work, we implicate potential applications of metabolic network reconstruction of *Euglena* through metabolic engineering. Overall, we anticipate that the use of model-assisted metabolic engineering in *Euglena* will increase in the near future.

Author contributions

SI and PS were responsible for the main writing of this manuscript. All authors contributed to the article and approved the submitted version.

Funding

This research work was partially supported by Chiang Mai University.

Acknowledgments

We thank Nick Kruger, George Ratcliffe and Ellis O'Neill, SI's DPhil supervisors who have help initiating the work related to this topic.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

References

Anfelt, J., Kaczmarzyk, D., Shabestary, K., Renberg, B., Rockberg, J., Nielsen, J., et al. (2015). Genetic and nutrient modulation of acetyl-CoA levels in *Synechocystis* for n-butanol production. *Microb. Cell Factories* 14:167. doi: 10.1186/s12934-015-0355-9

Barsanti, L., Bastianini, A., Passarelli, V., Tredici, M. R., and Gualtieri, P. (2000). Fatty acid content in wild type and WZSL mutant of *Euglena gracilis*. *J. Appl. Phycol.* 12, 515–520. doi: 10.1023/A:1008187514624

Becker, I., Prasad, B., Ntefidou, M., Daiker, V., Richter, P., and Lebert, M. (2021). Agrobacterium tumefaciens-Mediated Nuclear Transformation of a Biotechnologically Important Microalga—Euglena gracilis. *Int. J. Mol. Sci.* 22:6299. doi: 10.3390/ijms22126299

Chang, R. L., Ghamsari, L., Manichaikul, A., Hom, E. F., Balaji, S., Fu, W., et al. (2011). Metabolic network reconstruction of *Chlamydomonas* offers insight into light-driven algal metabolism. *Mol. Syst. Biol.* 7:518. doi: 10.1038/msb.2011.52

Chen, Z., Chen, Z., Zhu, J., He, J., Liu, Q., Zhu, H., et al. (2022). Proteomic responses of dark-adapted *Euglena gracilis* and bleached mutant against light stimuli. *Front. Bioeng. Biotechnol.* 10:843414. doi: 10.3389/fbioe.2022.843414

Cordoba, J., Perez, E., Van Vlierberghe, M., Bertrand, A. R., Lupo, V., Cardol, P., et al. (2021). De novo transcriptome meta-assembly of the mixotrophic freshwater microalga *Euglena gracilis*. *Genes* 12:842. doi: 10.3390/genes12060842

Dobáková, E., Flegontov, P., Skalický, T., and Lukeš, J. (2015). Unexpectedly streamlined mitochondrial genome of the Euglenozoan *Euglena gracilis*. *Genome Biol. Evol.* 7, 3358–3367. doi: 10.1093/gbe/evv229

Doetsch, N. A., Favreau, M. R., Kuscuoglu, N., Thompson, M. D., and Hallick, R. B. (2001). Chloroplast transformation in *Euglena gracilis*: splicing of a group III twintron transcribed from a transgenic psbK operon. *Curr. Genet.* 39, 49–60. doi: 10.1007/s002940000174

Ebenezer, T. E., Zoltner, M., Burrell, A., Nenarokova, A., Novák Vanclová, A. M. G., Prasad, B., et al. (2019). Transcriptome, proteome and draft genome of *Euglena gracilis*. *BMC Biol.* 17:11. doi: 10.1186/s12915-019-0626-8

El Khoury, D., Cuda, C., Luhovyy, B. L., and Anderson, G. H. (2012). Beta glucan: health benefits in obesity and metabolic syndrome. *J Nutr Metab* 2012:851362. doi: 10.1155/2012/851362

Fang, X., Lloyd, C. J., and Palsson, B. O. (2020). Reconstructing organisms in silico: genome-scale models and their emerging applications. *Nat. Rev.* Microbiology 18, 731–743. doi: 10.1038/s41579-020-00440-4

Fatma, Z., Hartman, H., Poolman, M. G., Fell, D. A., Srivastava, S., Shakeel, T., et al. (2018). Model-assisted metabolic engineering of *Escherichia coli* for long chain alkane and alcohol production. *Metabolic Engineering* 46, 1–12. doi: 10.1016/j.ymben.2018.01.002

Ferreira, R., Skrekas, C., Hedin, A., Sánchez, B. J., Siewers, V., Nielsen, J., et al. (2019). Model-assisted fine-tuning of central carbon metabolism in yeast through dCas9-based regulation. *ACS Synth. Biol.* 8, 2457–2463. doi: 10.1021/acssynbio.9b00258

Gissibl, A., Sun, A., Care, A., Nevalainen, H., and Sunna, A. (2019). Bioproducts from *Euglena gracilis*: synthesis and applications. *Front. Bioeng. Biotechnol.* 7:108. doi: 10.3389/fbioe.2019.00108

Griffiths, AJF, Wessler, SR, Carroll, SB, and Doebley, J. (2015). *Introduction to Genetic Analysis*, *11th*. New York, NY: W.H. Freeman & Company.

Hallick, R. B., Hong, L., Drager, R. G., Favreau, M. R., Monfort, A., Orsat, B., et al. (1993). Complete sequence of *Euglena gracilis* chloroplast DNA. *Nucleic Acids Res.* 21, 3537–3544. doi: 10.1093/nar/21.15.3537

Halter, D., Andres, J., Plewniak, F., Poulain, J., Da Silva, C., Arsène-Ploetze, F., et al. (2015). Arsenic hypertolerance in the protist Euglena mutabilis is mediated by specific

transporters and functional integrity maintenance mechanisms. *Environ. Microbiol.* 17, 1941–1949. doi: 10.1111/1462-2920.12474

Hammond, M. J., Nenarokova, A., Butenko, A., Zoltner, M., Dobáková, E. L., Field, M. C., et al. (2020). A uniquely complex mitochondrial proteome from *Euglena gracilis*. *Mol. Biol. Evol.* 37, 2173–2191. doi: 10.1093/molbev/msaa061

Hanna, E. M., Zhang, X., Eide, M., Fallahi, S., Furmanek, T., Yadetie, F., et al. (2020). ReCodLiver0.9: overcoming challenges in genome-scale metabolic reconstruction of a non-model species. *Front. Mol. Biosci.* 7:591406. doi: 10.3389/fmolb.2020.591406

Harada, R., Nomura, T., Yamada, K., Mochida, K., and Suzuki, K. (2020). Genetic engineering strategies for euglena gracilis and its industrial contribution to sustainable development goals: A review. *Front. Bioeng. Biotechnol.* 14:790. doi: 10.3389/fbioe.2020.00790

He, J., Liu, C., Du, M., Zhou, X., Hu, Z., Lei, A., et al. (2021). Metabolic responses of a model green microalga *Euglena gracilis* to different environmental stresses. *Front. Bioeng. Biotechnol.* 9:662655. doi: 10.3389/fbioe.2021.662655

Hirokawa, Y., Matsuo, S., Hamada, H., Matsuda, F., and Hanai, T. (2017). Metabolic engineering of *Synechococcus* elongatus PCC 7942 for improvement of 1,3-propanediol and glycerol production based on in silico simulation of metabolic flux distribution. *Microb. Cell Factories* 16:212. doi: 10.1186/s12934-017-0824-4

Inui, H., Miyatake, K., Nakano, Y., and Kitaoka, S. (1982). Wax ester fermentation in *Euglena gracilis*. *Fed. Eur. Biochem. Soc. Lett.* 150, 89–93. doi: 10.1016/0014-5793(82)81310-0

Inwongwan, S. (2021). *Metabolic Flux Analysis in Euglena gracilis*. University of Oxford.

Inwongwan, S., Kruger, N. J., Ratcliffe, R. G., and O'Neill, E. C. (2019). Euglena central metabolic pathways and their subcellular locations. *Meta* 9:115. doi: 10.3390/metabo9060115

Ishikawa, M., Nomura, T., Tamaki, S., Ozasa, K., Suzuki, T., Toyooka, K., et al. (2022). RISPR/Cas9-mediated generation of non-motile mutants to improve the harvesting efficiency of mass-cultivated Euglena gracilis. *Plant. Biotechnol. J.* 20, 2042–2044. doi: 10.1111/pbi.13904

Kato, S., Soshino, M., Takaichi, S., Ishikawa, T., Nagata, N., Asahina, M., et al. (2017). Suppression of the phytoene synthase gene (EgcrtB) alters carotenoid content and intracellular structure of *Euglena gracilis*. *BMC Plant Biol.* 17:125. doi: 10.1186/s12870-017-1066-7

Khatiwada, B., Hasan, M. T., Sun, A., Kamath, K. S., Mirzaei, M., Sunna, A., et al. (2020). Proteomic response of *Euglena gracilis* to heavy metal exposure – identification of key proteins involved in heavy metal tolerance and accumulation. *Algal Res.* 45:101764. doi: 10.1016/j.algal.2019.101764

Khatiwada, B., Kautto, L., Sunna, A., Sun, A., and Nevalainen, H. (2019). Nuclear transformation of the versatile microalga *Euglena gracilis*. *Algal Res.* 37, 178–185. doi: 10.1016/j.algal.2018.11.022

Kimura, M., and Ishikawa, T. (2018). Suppression of DYRK ortholog expression affects wax ester fermentation in *Euglena gracilis*. *J. Appl. Phycol.* 30, 367–373. doi: 10.1007/s10811-017-1235-y

Kliphuis, A. M. J., Klok, A. J., Martens, D. E., Lamers, P. P., Janssen, M., and Wijffels, R. H. (2012). Metabolic modeling of *Chlamydomonas reinhardtii*: energy requirements for photoautotrophic growth and maintenance. *J. Appl. Phycol.* 24, 253–266. doi: 10.1007/s10811-011-9674-3

Korn, E. D. (1964). The fatty acids of *Euglena gracilis*. *J. Lipid Res.* 5, 352–362. doi: 10.1016/S0022-2275(20)40204-4

Kruger, N. J., and Ratcliffe, R. G. (2015). Fluxes through plant metabolic networks: measurements, predictions, insights and challenges. *Biochem. J.* 465, 27–38. doi: 10.1042/BJ20140984

Kusmic, C., Barsacchi, R., Barsanti, L., Gualtieri, P., and Passarelli, V. (1998). *Euglena gracilis* as source of the antioxidant vitamin E. effects of culture conditions in the wild strain and in the natural mutant WZSL. *J. Appl. Phycol.* 10, 555–559. doi: 10.1023/A:1008022305865

Lerman, J. A., Hyduke, D. R., Latif, H., Portnoy, V. A., Lewis, N. E., Orth, J. D., et al. (2012). In silico method for modelling metabolism and gene product expression at genome scale. *Nat. Commun.* 3:929. doi: 10.1038/ncomms1928

Li, C.-T., Yelsky, J., Chen, Y., Zuñiga, C., Eng, R., Jiang, L., et al. (2019). Utilizing genome-scale models to optimize nutrient supply for sustained algal growth and lipid productivity. *NPJ Syst. Biol. Appl.* 5:33. doi: 10.1038/s41540-019-0110-7

Liu, X., Ding, W., and Jiang, H. (2017). Engineering microbial cell factories for the production of plant natural products: from design principles to industrial-scale production. *Microb. Cell Factories* 16:125. doi: 10.1186/s12934-017-0732-7

Liu, Y., Khusnutdinova, A., Chen, J., Crisante, D., Batyrova, K., Raj, K., et al. (2022). Systems engineering of *Escherichia coli* for n-butane production. *Metab. Eng.* 74, 98–107. doi: 10.1016/j.ymben.2022.10.001

Mangal, V., Donaldson, M. E., Lewis, A., Saville, B. J., and Guéguen, C. (2022). Identifying *Euglena gracilis* metabolic and transcriptomic adaptations in response to mercury stress. *Front. Environ. Sci.* 10:836732. doi: 10.3389/fenvs.2022.836732

Marmiesse, L., Peyraud, R., and Cottret, L. (2015). FlexFlux: combining metabolic flux and regulatory network analyses. *BMC Syst. Biol.* 9:93. doi: 10.1186/s12918-015-0238-z

Murphy, E. J., Rezoagli, E., Major, I., Rowan, N. J., and Laffey, J. G. (2020). β-Glucan metabolic and immunomodulatory properties and potential for clinical application. *J Fungi (Basel)* 6:356. doi: 10.3390/jof6040356

Nakazawa, M., Andoh, H., Koyama, K., Watanabe, Y., Nakai, T., Ueda, M., et al. (2015). Alteration of wax Ester content and composition in *Euglena gracilis* with gene silencing of 3-ketoacyl-CoA Thiolase isozymes. *Lipids* 50, 483–492. doi: 10.1007/s11745-015-4010-3

Nomura, T., Yoshikawa, M., Suzuki, K., and Mochida, K. (2020). Highly Efficient CRISPR-Associated Protein 9 Ribonucleoprotein-Based Genome Editing in Euglena gracilis. *STAR Protoc.* 1:100023. doi: 10.1016/j.xpro.2020.100023

Novák Vanclová, A. M. G., Zoltner, M., Kelly, S., Soukal, P., Záhonová, K., Füssy, Z., et al. (2020). Metabolic quirks and the colourful history of the *Euglena gracilis* secondary plastid. *New Phytol.* 225, 1578–1592. doi: 10.1111/nph.16237

O'Neill, E. C., Trick, M., Henrissat, B., and Field, R. A. (2015b). Euglena in time: evolution, control of central metabolic processes and multi-domain proteins in carbohydrate and natural product biochemistry. *Perspect. Sci.* 6, 84–93. doi: 10.1016/j.pisc.2015.07.002

O'Neill, E. C., Trick, M., Hill, L., Rejzek, M., Dusi, R. G., Hamilton, C. J., et al. (2015a). The transcriptome of *Euglena gracilis* reveals unexpected metabolic capabilities for carbohydrate and natural product biochemistry. *Mol. BioSyst.* 11, 2808–2820. doi: 10.1039/C5MB00319A

Ogbonna, J. C., Tomiyama, S., and Tanaka, H. (1998). Heterotrophic cultivation of *Euglena gracilis* Z for efficient production of alpha-tocopherol. *J. Appl. Phycol.* 10, 67–74. doi: 10.1023/A:1008011201437

Ohmachi, M., Fujiwara, Y., Muramatsu, S., Yamada, K., Iwata, O., Suzuki, K., et al. (2016). A modified single-cell electroporation method for molecule delivery into a motile protist, *Euglena gracilis*. *J. Microbiol. Methods* 130, 106–111. doi: 10.1016/j.mimet.2016.08.018

Passi, A., Tibocha-Bonilla, J. D., Kumar, M., Tec-Campos, D., Zengler, K., and Zuniga, C. (2021). Genome-scale metabolic modeling enables in-depth understanding of big data. *Meta* 12:14. doi: 10.3390/metabo12010014

Poolman, M. G., Miguet, L., Sweetlove, L. J., and Fell, D. A. (2009). A genome-scale metabolic model of Arabidopsis and some of its properties. *Plant Physiol.* 151, 1570–1581. doi: 10.1104/pp.109.141267

Prigent, S., Frioux, C., Dittami, S. M., Thiele, S., Larhlimi, A., Collet, G., et al. (2017). Meneco, a topology-based gap-filling tool applicable to degraded genome-wide metabolic networks. *PLoS Comput. Biol.* 13:e1005276. doi: 10.1371/journal.pcbi.1005276

Santos-Merino, M., Singh, A. K., and Ducat, D. C. (2019). New applications of synthetic biology tools for cyanobacterial metabolic engineering. *Front. Bioeng. Biotechnol.* 7:33. doi: 10.3389/fbioe.2019.00033

Takeyama, H., Kanamaru, A., Yoshino, Y., Kakuta, H., Kawamura, Y., and Matsunaga, T. (1997). Production of antioxidant vitamins, beta-carotene, vitamin C, and vitamin E, by two-step culture of *Euglena gracilis* Z. *Biotechnol. Bioeng.* 53, 185–190. doi: 10.1002/(SICI)1097-0290(19970120)53:2<185::AID-BIT8>3.0.CO;2-K

Tamaki, S., Nishino, K., Ogawa, T., Maruta, T., Sawa, Y., Arakawa, K., et al. (2020). Comparative proteomic analysis of mitochondria isolated from *Euglena gracilis* under aerobic and hypoxic conditions. *PLoS One* 14:e0227226. doi: 10.1371/journal.pone.0227226

Thiele, I., and Palsson, B. Ø. (2010). A protocol for generating a high-quality genome-scale metabolic reconstruction. *Nat. Protoc.* 5, 93–121. doi: 10.1038/nprot.2009.203

Tibocha-Bonilla, J. D., Zuñiga, C., Godoy-Silva, R. D., and Zengler, K. (2018). Advances in metabolic modeling of oleaginous microalgae. *Biotechnol. Biofuels* 11:241. doi: 10.1186/s13068-018-1244-3

Wang, X., Liu, W., Xin, C., Zheng, Y., Cheng, Y., Sun, S., et al. (2016). Enhanced limonene production in cyanobacteria reveals photosynthesis limitations. *Proc. Natl. Acad. Sci.* 113, 14225–14230. doi: 10.1073/pnas.1613340113

Yamane, Y.-I., Utsunomiya, T., Watanabe, M., and Sasaki, K. (2001). Biomass production in mixotrophic culture of *Euglena gracilis* under acidic condition and its growth energetics. *Biotechnol. Lett.* 23, 1223–1228. doi: 10.1023/A:1010573218863

Yan, Q., and Fong, S. S. (2017). Challenges and advances for genetic engineering of non-model bacteria and uses in consolidated bioprocessing. *Front. Microbiol.* 8:2060. doi: 10.3389/fmicb.2017.02060

Yoshida, Y., Tomiyama, T., Maruta, T., Tomita, M., Ishikawa, T., and Arakawa, K. (2016). De novo assembly and comparative transcriptome analysis of *Euglena gracilis* in response to anaerobic conditions. *BMC Genomics* 17:182. doi: 10.1186/s12864-016-2540-6

Yoshikawa, K., Toya, Y., and Shimizu, H. (2017). Metabolic engineering of *Synechocystis* sp. PCC 6803 for enhanced ethanol production based on flux balance analysis. *Bioprocess Biosyst. Eng.* 40, 791–796. doi: 10.1007/s00449-017-1744-8

Yuan, H., Cheung, C. Y. M., Poolman, M. G., Hilbers, P. A. J., and Van Riel, N. A. W. (2016). A genome-scale metabolic network reconstruction of tomato (*Solanum lycopersicum* L.) and its application to photorespiratory metabolism. *Plant J.* 85, 289–304. doi: 10.1111/tpj.13075

Zuñiga, C., Li, C.-T., Huelsman, T., Levering, J., Zielinski, D. C., Mcconnell, B. O., et al. (2016). Genome-scale metabolic model for the green alga *Chlorella vulgaris* UTEX 395 accurately predicts phenotypes under autotrophic, heterotrophic, and mixotrophic growth conditions. *Plant Physiol.* 172, 589–602. doi: 10.1104/pp.16.00593

frontiers | Frontiers in Microbiology

TYPE Original Research
PUBLISHED 14 March 2023
DOI 10.3389/fmicb.2023.1126030

OPEN ACCESS

EDITED BY
Xiaoming Tan,
Hubei University,
China

REVIEWED BY
Weiwen Zhang,
Tianjin University,
China
Fei Gan,
University of California,
Berkeley,
United States

*CORRESPONDENCE
María Santos-Merino
✉ santosm7@msu.edu
Juan Nogales
✉ j.nogales@csic.es

†PRESENT ADDRESS
María Santos-Merino,
MSU-DOE Plant Research Laboratory,
Michigan State University,
East Lansing,
MI, United States

SPECIALTY SECTION
This article was submitted to
Microbial Physiology and Metabolism,
a section of the journal
Frontiers in Microbiology

RECEIVED 10 January 2023
ACCEPTED 22 February 2023
PUBLISHED 14 March 2023

CITATION
Santos-Merino M, Gargantilla-Becerra Á, de la Cruz F and Nogales J (2023) Highlighting the potential of *Synechococcus elongatus* PCC 7942 as platform to produce α-linolenic acid through an updated genome-scale metabolic modeling.
Front. Microbiol. 14:1126030.
doi: 10.3389/fmicb.2023.1126030

Highlighting the potential of *Synechococcus elongatus* PCC 7942 as platform to produce α-linolenic acid through an updated genome-scale metabolic modeling

María Santos-Merino[1*†], Álvaro Gargantilla-Becerra[2,3], Fernando de la Cruz[1] and Juan Nogales[2,3*]

[1]Instituto de Biomedicina y Biotecnología de Cantabria, Universidad de Cantabria—CSIC, Santander, Cantabria, Spain, [2]Department of Systems Biology, Centro Nacional de Biotecnología (CSIC), Madrid, Spain, [3]Interdisciplinary Platform for Sustainable Plastics towards a Circular Economy-Spanish National Research Council (SusPlast-CSIC), Madrid, Spain

Cyanobacteria are prokaryotic organisms that capture energy from sunlight using oxygenic photosynthesis and transform CO_2 into products of interest such as fatty acids. *Synechococcus elongatus* PCC 7942 is a model cyanobacterium efficiently engineered to accumulate high levels of omega-3 fatty acids. However, its exploitation as a microbial cell factory requires a better knowledge of its metabolism, which can be approached by using systems biology tools. To fulfill this objective, we worked out an updated, more comprehensive, and functional genome-scale model of this freshwater cyanobacterium, which was termed *i*MS837. The model includes 837 genes, 887 reactions, and 801 metabolites. When compared with previous models of *S. elongatus* PCC 7942, *i*MS837 is more complete in key physiological and biotechnologically relevant metabolic hubs, such as fatty acid biosynthesis, oxidative phosphorylation, photosynthesis, and transport, among others. *i*MS837 shows high accuracy when predicting growth performance and gene essentiality. The validated model was further used as a test-bed for the assessment of suitable metabolic engineering strategies, yielding superior production of non-native omega-3 fatty acids such as α-linolenic acid (ALA). As previously reported, the computational analysis demonstrated that *fabF* overexpression is a feasible metabolic target to increase ALA production, whereas deletion and overexpression of *fabH* cannot be used for this purpose. Flux scanning based on enforced objective flux, a strain-design algorithm, allowed us to identify not only previously known gene overexpression targets that improve fatty acid synthesis, such as Acetyl-CoA carboxylase and β-ketoacyl-ACP synthase I, but also novel potential targets that might lead to higher ALA yields. Systematic sampling of the metabolic space contained in *i*MS837 identified a set of ten additional knockout metabolic targets that resulted in higher ALA productions. *In silico* simulations under photomixotrophic conditions with acetate or glucose as a carbon source boosted ALA production levels, indicating that photomixotrophic nutritional regimens could be potentially exploited *in vivo* to improve fatty acid production in cyanobacteria. Overall, we show that *i*MS837 is a powerful computational platform that proposes new metabolic engineering strategies to produce biotechnologically relevant compounds, using *S. elongatus* PCC 7942 as non-conventional microbial cell factory.

KEYWORDS

cyanobacteria, *Synechococcus elongatus* PCC 7942, genome-scale metabolic model, strain-designing algorithms, α-linolenic acid

1. Introduction

Cyanobacteria are promising host organisms for the production of compounds with biotechnological applications (Santos-Merino et al., 2019). Their ability to utilize solar energy to fix CO_2 makes them particularly attractive, especially in an era where the urge of development of sustainable biotechnological processes has gained an increased attention (Rajneesh et al., 2017). As bioproduction platforms, cyanobacteria offer several advantages when compared to plants and algae, such as higher photosynthetic efficiencies (Zahra et al., 2020) and ease of genetic manipulation (Berla et al., 2013). The model cyanobacterium *Synechococcus elongatus* PCC 7942 has been widely explored as a cell factory to produce several value-added compounds, including 2,3-butanediol (Oliver et al., 2013; Nozzi and Atsumi, 2015) and omega-3 fatty acids (Santos-Merino et al., 2018, 2022), among others.

Cyanobacteria are able to naturally produce short-chain omega-3 fatty acids, such as alpha-linolenic acid (ALA) and stearidonic acid (SDA). In the quest to find more sustainable, suitable, and economically viable hosts for the production of omega-3 fatty acids, cyanobacteria, and microalgae have emerged as alternative organisms to native (i.e., fish and plant oils and oleaginous microorganisms) and non-native sources (e.g., genetic engineering organisms) (Galán et al., 2019; Patel et al., 2020). By contrast, cyanobacteria are preferred organisms over microalgae due to their small genomes that generally facilitates manipulation and the availability of a large number of advanced genome editing tools for cyanobacterial genetic engineering (Vavitsas et al., 2021). In the last years, extensive research efforts have been focused on the metabolic engineering of cyanobacterial strains to enhance ALA and SDA production (Dong et al., 2016; Yoshino et al., 2017; Santos-Merino et al., 2018; Poole et al., 2020; Santos-Merino et al., 2022). In many of the cases, the enzymes directly involved in the synthesis of omega-3 fatty acids (i.e., desaturases) have been overexpressed to increase the production yields. There are only a couple of reports where other targets have been exploited to increase omega-3 fatty acids, such as enzymes involved in the saturated fatty acid synthesis (Santos-Merino et al., 2018, 2022) and the vesicle-inducing protein in plastids (Vipp1), a thylakoid membrane formation enhancer (Poole et al., 2020). Limited exploration has been done to identify additional targets in other competitive metabolic pathways with the aim to increase omega-3 fatty acid titers. One of the major obstacles to make omega-3 fatty acid production by engineered cyanobacteria practical and cost-effective is the low productivity levels achieved in these engineered strains (Shinde et al., 2022). Traditional metabolic engineering strategies are the most common avenues used to increase production yields in cyanobacteria. However, since these classical technologies are expensive, time-consuming, and labor-intensive processes, computational biology strategies are emerging as powerful tools to overcome these limitations (Xu C. et al., 2013; Gudmundsson and Nogales, 2021).

Genome-scale models (GEMs) are based on the annotated genome sequence and describe metabolic pathways as stoichiometric coefficients and mass balances of participating metabolites (Gudmundsson and Nogales, 2015). They can be used as computational test-bed to estimate metabolic fluxes using numerical optimization, thus offering a systems-biology tool not only to link genotype to phenotype but also to analyze and contextualize the metabolic capabilities of organisms (Oberhardt et al., 2009). GEMs have been successfully applied to analyze and guide the metabolism of cyanobacteria for production of several target compounds from CO_2 (Nogales et al., 2013; Santos-Merino et al., 2019; Hendry et al., 2020). To date, three GEMs have been developed for *S. elongatus* PCC 7942: *i*Syf715 (Triana et al., 2014), *i*JB785 (Broddrick et al., 2016), and *i*JB792 (Broddrick et al., 2019). However, these existing models have paid little attention to fatty acid biosynthetic pathways, making it harder to use them as tool to analyze the potential of *S. elongatus* PCC 7942 toward the production of omega-3 fatty acids and related compounds. Then, an updated GEM with a high-quality annotation of fatty acid biosynthetic pathways is urgently required for accurately contextualizing fatty acid metabolism while predicting nutritional, physiological, and genetic scenarios for overproducing omega-3 fatty acids.

Full-facing this challenge, we present here *i*MS837, an updated GEM of *S. elongatus* PCC 7942 using *i*JB792 as a foundation (Broddrick et al., 2019). We validate the accuracy of our GEM using growth performance and gene essentiality predictions. The validated GEM was subsequently used to assess the production of omega-3 fatty, identifying the overexpression of *fabF* as key metabolic target to increase ALA production in agreement with published experimental data (Santos-Merino et al., 2018). Next, we used the model to predict possible engineered metabolic targets to enhance ALA production, identifying a set of ten additional knock-out metabolic targets that resulted in higher yields of this omega-3 fatty acid, albeit not growth-coupled. In addition, mixotrophic conditions using different carbon sources (i.e., glucose and acetate) were evaluated *in silico*, boosting the ALA yields obtained in both cases. These predictions will serve as a starting point for future efforts to design strains and conditions that will potentially improve omega-3 fatty acid production in cyanobacteria.

2. Results

2.1. Properties of the *i*MS837 GEM metabolic network

*i*MS837 was constructed using as a template the previously published GEM, *i*JB792 (Broddrick et al., 2019). We expanded *i*JB792 by adding new content, including 46 genes, 24 metabolites, and 29 reactions by means of a detailed manual curation based on literature legacy and by comparison with other published high-quality GEMs

from cyanobacteria [i.e., *i*JN678 from *Synechocystis*, sp. PCC 6803 (Nogales et al., 2012)] and from heterotrophic bacteria [i.e., *i*JN1462 from *Pseudomonas putida* KT2440 (Nogales et al., 2020)] (Table 1; Supplementary Table S1). Overall, the largest portion of the new metabolic content of *i*MS837 was related with the fatty acid metabolism including 22 new reactions related to this subsystem (Figure 1A). In addition, other 11 reactions were included in *i*MS837, whereas four reactions related to the intracellular demands were removed. These new reactions demonstrate the uniqueness of *i*MS837, especially regarding the modeling of the fatty acid biosynthesis.

Beyond the important metabolic expansion done, the gene–protein-reaction (GPR) associations of several reactions were updated while others were corrected in order to improve the model accuracy (Supplementary Dataset 1). For example, reactions involved in glycogen synthesis and degradation were corrected while reactions related to electron transport chain (e.g., Cytochrome b_6f complex, Cytochrome aa3 oxidase, Ferredoxin:NADPH oxidoreductase) were updated, among others. In addition, a meticulous analysis of orphan reactions included in *i*JB792 was carried out in order to identify the genes responsible for such reactions. We manually added 20 new GPRs, which were mainly involved in photosynthesis, respiration, and photorespiration processes (Figure 1B). The rest of the added genes (i.e., 26) belong to a large variety of subsystems as shown in Figure 1B. Finally, the biomass objective function (BOF) of *i*JB792 was also updated in *i*MS837 by removing those metabolites that were not necessary for growth (Supplementary Dataset 1). The modification of the BOF had associated changes in: (i) pigments and xanthophylls, (ii) cofactor pools, and (iii) lipids.

MEMOTE is a platform that has been developed to promote standardization of GEMs, as well as to assess quality control metrics in order to improve model reproducibility and applicability (Lieven et al., 2020). Therefore, we used MEMOTE tool in order to define the completeness, consistency, and interoperability of *i*MS837 when compared with previous models while analyzing potential flaws or shortcomings (Supplementary Dataset 2). The overall score for the model was 75% over the 20% estimated for *i*JB792, which suggest a very good level of completeness. The scores in annotation subcategories were increased by adding annotations and Systems Biology Ontology (SBO) terms to metabolites, reactions, and genes in the updated GEM which were not previously included in *i*JB792. The model scored 55% for the critical category of consistency, which represents accuracy in reaction stoichiometry, mass and charge balances, connectivity of metabolites, and reaction cycles. A major gap was found due the lack of annotation of outside references for some genes, metabolites, and reactions. This limitation only will have some impact when using automated tools or scripts; however, its accuracy and usability should not be affected. Taking together, the MEMOTE analysis demonstrated that *i*MS837 is a highly complete and detailed model that can be used as a reference for other GEM constructions.

TABLE 1 Properties of the different GEM models of *S. elongatus* PCC 7942.

	Values for			
	*i*MS837	*i*JB792	*i*JB785	*i*Syf715
Metabolites	801	777	768	838
Reactions	887	858	850	851
Genes	837	791	785	715
Reference	This work	Broddrick et al. (2019)	Broddrick et al. (2016)	Triana et al. (2014)

2.2. Model validation using gene essentiality prediction

An increase in the number of genes, reactions and metabolites does not always indicate a higher-quality GEM. In order to validate the quality of *i*MS837, we conducted an extensive gene essentiality analysis of the genes included in this GEM by comparing the predicted results *in silico* with available essentiality experimental data for *S. elongatus* PCC 7942 (Rubin et al., 2015; Figure 2). We performed single-gene knockout simulations in *i*MS837 using COBRApy (Ebrahim et al., 2013). The model-based gene essentiality predictions showed an overall high-level of accuracy, 85.5% (Figure 2B). More specifically, the model was able to correctly assign 330 and 333 as essential and non-essential genes, respectively (Figure 2A; Supplementary Table S2). On the contrary, the level of discrepancy found between the predictions of *i*MS837 and the experimental data was pretty low. For instance, 75 genes were incorrectly predicted as essential while only 37 genes were predicted as essential but were found non-essential *in vivo*. Gene essentiality accuracy assignment in GEMs is often biased due large number of non-essential genes (Wei et al., 2013). Therefore, to avoid potential bias caused by such effect, we proceed to additionally compute the sensitivity (i.e., proportion of essential genes that have been correctly identified), specificity (i.e., proportion of true negatives that have been correctly predicted), and precision (i.e., the probability that the essential genes were predicted as essential) of our gene essentially prediction (Figure 2B). We found very high values for all these parameters, suggesting no significant bias, and corroborating the high capacity of *i*MS837 when predicting gene essentiality.

On the other hand, when comparing the level of accuracy of *i*MS837 with that from previous *S. elongatus* PCC 7942 GEMs, *i*JB792 (Broddrick et al., 2019), *i*JB785 (Broddrick et al., 2016), and *i*Syf715 (Triana et al., 2014), we found that *i*MS837 scored the highest in terms of accuracy, sensitivity, specificity, and precision (Figure 2B). Whereas *i*JB792, *i*JB785, and *i*Syf715 were able to correctly assign the 80.3, 79.2, and 60.1% of the genes included in the model as true essential and nonessential, the 80.3, 79.2, and 60.1% of the genes included in the model, *i*MS837 outperforms better that these three GEMs by correctly predicting 85.5% (Figure 2B; Supplementary Table S2). Overall, these results indicate that we have not only expanded the last published GEM of *S. elongatus* PCC 7942 by adding new reactions, metabolites and genes, but also, we have improved its accuracy by increasing its completeness.

2.3. System evaluation of *S. elongatus* PCC 7942 as a cell factory toward the production of omega-3 fatty acids

The capability of a GEM to provide accurate predictions of experimentally supported data of a target organism's functional states is a key feature in order to assess the accuracy and completeness of the final reconstruction. Once *i*MS837's accuracy and completeness were assessed, the model was ready to be used to characterize metabolic

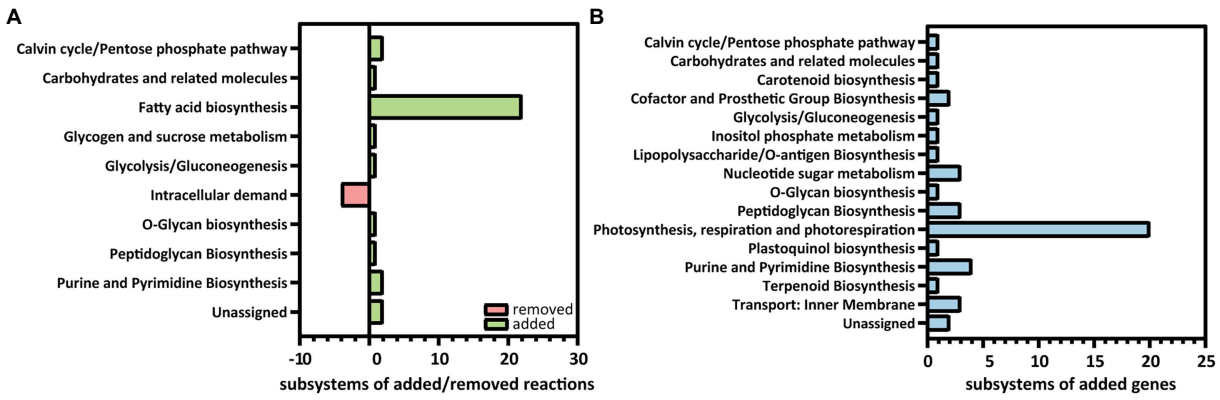

FIGURE 1
Distribution of reactions and genes added and removed to *i*MS837 based on functional subsystems. **(A)** Summary of reactions added and removed to *i*MS837. **(B)** Summary of genes added to *i*MS837. In both cases, the subsystems where changes have been made are only represented.

A

		in silico	
		Essential	Non-essential
in vivo	Essential	True Positive 330 (41.3%)	False Positive 37 (9.4%)
	Non-essential	False Negative 75 (7.8%)	True Negative 333 (41.6%)

B

Parameters for evaluating gene essentiality predictions						
GEM	Accuracy	Sensitivity	Specificity	Precision	F1-score	Kappa
*i*MS837	85.5%	81.5%	90%	89.9%	0.855	0.712
*i*JB792	80.3%	78%	82.3%	79.6%	0.788	0.604
*i*JB785	79.2%	87.7%	86.3%	88.8%	0.883	0.719
*i*Syf715	60.1%	32.8%	96.7%	93.7%	0.486	0.267

FIGURE 2
Comparison of gene essentially predictions to experimental results. **(A)** Contingency table of the results obtained for gene essentially predictions using *i*MS837. **(B)** Parameters used to estimate the performance of gene essentially predictions for all GEM of *Synechococcus elongatus* PCC 7942. Gene essentially results obtained *in silico* revealed a high level of accuracy (85.5%), sensitivity (81.5%), specificity (90%) and precision (89.9%) of *i*MS837. Genes with ambiguous essentiality results and not-analyzed *in vivo* were excluded from this analysis (Supplementary Table S2).

states underlying observed phenotypic functions and as a computational framework for metabolic engineering endeavors. In this regard, we further used *i*MS837 as a test-bed to systematically analyze *S. elongatus* PCC 7942 as a cell factory toward the production of ALA. In order to do that, we first included the non-native ALA synthesis pathways in the GEM of *S. elongatus* PCC 7942, a non-natural ALA producer strain. We used *i*JN678, a well-developed GEM from *Synechocystis* sp. PCC 6803 (Nogales et al., 2012) that includes this biosynthetic pathway, since this cyanobacterium is able to naturally produce omega-3 fatty acids (Tasaka et al., 1996). A total of 34 reactions and 28 metabolites were added to *i*MS837 in order to produce ALA and accumulate this omega-3 fatty acid in its membranes (Supplementary Table S3; Supplementary Dataset 1). Among the added reactions to produce ALA in *S. elongatus* PCC 7942, we included the desaturases DesA and DesB (Δ12- and Δ15-desaturases) and all the reactions involved in integrating this omega-3 in the phospholipids of cellular membranes.

In silico, the production capabilities of a given strain can be shown using production envelope plots, which represent all possible production rates of a selected metabolite and their associated feasible growth rates (Lewis et al., 2012). Production envelope for ALA production showed that, although feasible under a real nutritional scenario, the production of this omega-3 fatty acid is not coupled to growth (Supplementary Figure S1). In other words, for all the possible levels of ALA production, the model predicted a decrease in the maximal growth rate. This is not surprising since the synthesis of fatty acids is one of the most energetically expensive process among all the lipid membranes components (Zhang and Rock, 2009), and the heterologous production of ALA directly compete with cell growth (Chen et al., 2014). Cyanobacteria only produced unsaturated fatty acids in respond to drops in temperature to compensate for the decrease in membrane fluidity, conditions where they do not normally growth. In addition, it is well-established that cyanobacteria are only able to accumulate fatty acids at low temperatures, where the transcripts for *desA* and *desB* desaturases are more abundant (Sakamoto et al., 1997; Ludwig and Bryant, 2012).

Once demonstrated that the model was able to predict ALA production fluxes, we decided to explore *in silico* previously identified

genetic interventions that led to a decrease and an increase in ALA yield (i.e., *fabH* deletion and *fabF* overexpression, respectively) (Santos-Merino et al., 2018). We used Markov chain Monte Carlo sampling to establish potential differences in the metabolic states between strains by comparing the allowed specific metabolic solution spaces (Schellenberger and Palsson, 2009). This flux sampling methodology allowed us to explore the feasible flux solutions in our metabolic network by generating probability distributions of steady-state reaction fluxes (Herrmann et al., 2019). We analyzed the fluxes of the reactions involved in the saturated and unsaturated fatty acid synthesis and their probability in each strain.

Firstly, we explored the flux distributions for strains with increased levels of FabF enzyme (3OAS180 reaction in *i*MS837): (i) FabF-UP-2x, with double flux for FabF reaction than the control strain; and (ii) FabF-UP-4x, with quadruple flux for FabF reaction than the control strain (Figure 3). We observed an overall increase in the flux of reactions involved in the saturated fatty acid synthesis pathway, proportional to the value of FabF flux. Same effect was observed in the flux of the DesC desaturase (DESAT18a). In addition, a large flux for DesA and DesB desaturases (DES::12 and DES::15, respectively) was observed for FabF-UP-4x, but the probability was lower than the minor fluxes observed FabF-UP-2x. These results agree with *in vivo* experimental data linking the overexpression of *fabF* with increased levels of C18:1 (the product of DESAT18a reaction) and ALA (Santos-Merino et al., 2018).

As a second scenario, we analyzed *in silico* the flux distributions for strains harboring down- and up-regulation of FabH fluxes (KAS15 reaction in *i*MS837) that we denominated FabH-DOWN and FabH-UP, respectively (Figure 4). For the FabH-UP sampling, it was predicted an increase in the flux of the reactions involved in the elongation cycle (FabF, FabG, FabZ, and FabI). As could be expected, we computed an increase of these fluxes for FabH-DOWN sampling. On the other hand, the flux of DesC desaturase (DESAT18a) was increased in FabH-UP, while it was almost zero for FabH-DOWN. Finally, FabH-UP and FabH-DOWN failed to increase the fluxes through DesA and DesB desaturases (DES::12 and DES::15, respectively). Overall, these results mainly agree with *in vivo* observations, showing no increased in ALA yields after modifications of the expression of *fabH* (Santos-Merino et al., 2018).

2.4. Identification of potential gene overexpression targets to increase ALA yields

GEM gives us the advantage of speeding up the exploration and redesign of the metabolism of an organism toward the production of a given metabolite. In most cases, the upregulation of certain fluxes directly or indirectly involved in the biosynthetic pathway of a desired compound, results in improved yields. With the aim of identifying fluxes that could be upregulated to increase ALA production, we used flux scanning based on enforced objective function (FSEOF) (Park et al., 2012). This method scans changes in metabolic fluxes in response to an artificially enforced objective flux of the desired product formation. Using this algorithm, up to 70 target reactions were identified as a result of gradually increments in the ALA production reaction and the acceptance of up to 20% reduction in the biomass-producing reaction (Supplementary Table S4).

The FSEOF simulation results revealed that ALA production increased with the enhancement of the fatty acid biosynthetic pathways (Figure 5). All the reactions involved in the unsaturation and desaturation steps in the fatty acid biosynthesis were identified as potential overexpression targets. Following the logic of increasing the intermediates of ALA synthesis, we have previously overproduced *in vivo* FabF, FabH, FabD, FabZ, and FabG with the aim to improve ALA yields (Santos-Merino et al., 2018, 2022), with the exception of FabI. The overexpression of most of them did not increase C18:1 levels, the substrate for the sequential activity of DesA and DesB desaturases. Only the overproduction of FabF was able to successfully increase ALA yields. Little known of the regulation of fatty acid synthesis in cyanobacteria, that could affect to our experimental interventions in this pathway, as well as the *in silico* predictive capability of *i*MS837, which does not include information about regulation. Acetyl-CoA carboxylase, which catalyzes the first step of the saturated fatty acid biosynthesis, was also identified as a potential overexpression target to increase ALA production. Overexpression of this enzyme has been proven to be an effective way to increase the rate of saturated fatty acid synthesis in *Synechocystis* sp. PCC 6803 (Eungrasamee et al., 2019), as well as other Acetyl-CoA derived compounds, such as alkanes and alkenes (Tan et al., 2011; Wang et al., 2013). All together indicates that overexpression of Acetyl-CoA could be a feasible strategy to be implemented *in vivo* with the aim to increase ALA yields.

In addition, the increase in the fluxes through the components of the photosynthetic electron transport chain was the second category of reactions that showed an increase in ALA production. It has been experimentally demonstrated that an increase in the light intensity leads to an increase in omega-3 fatty acids in *Synechococcus* sp. PCC 7002 (Sakamoto et al., 1997). Then, it is not surprising that reactions related to photosynthesis are the second-most represented in the results obtained with FSEOF. The third category of reactions that can be upregulated to increase ALA yields is the amino acid metabolism (Figure 5), including reactions involved in the synthesis of alanine, serine, glutamate, glutamine, and aspartate (Supplementary Table S4). In *Arthrospira platensis*, the supplementation of cultures with aspartate stimulated the accumulation of saturated fatty acids, possibly through enhanced *de novo* fatty acid biosynthesis (Fekrat et al., 2022). An increase of the *in silico* flux of reactions associated with amino acid metabolism could have the same role. Finally, it is important to highlight that in the category of reactions related to pyruvate metabolism and tricarboxylic acid cycle (TCA), the only reaction which flux should be increased is the one for the Acetyl-CoA synthetase (Supplementary Table S4). Acetyl-CoA is a key metabolic intermediate that links many metabolic processes, including the TCA cycle, amino acid metabolism, and fatty acid metabolism (Mills et al., 2020). It is possible that in order to increase the direct flux of Acetyl-CoA into the fatty acid synthesis, the flux of this key intermediate should be increased to avoid the competition between all the pathways that use it.

2.5. Identification of potential genetic interventions to improve ALA production

In addition to upregulation of reaction fluxes, gene knockout is one the most common strategies to improve microbial strains for

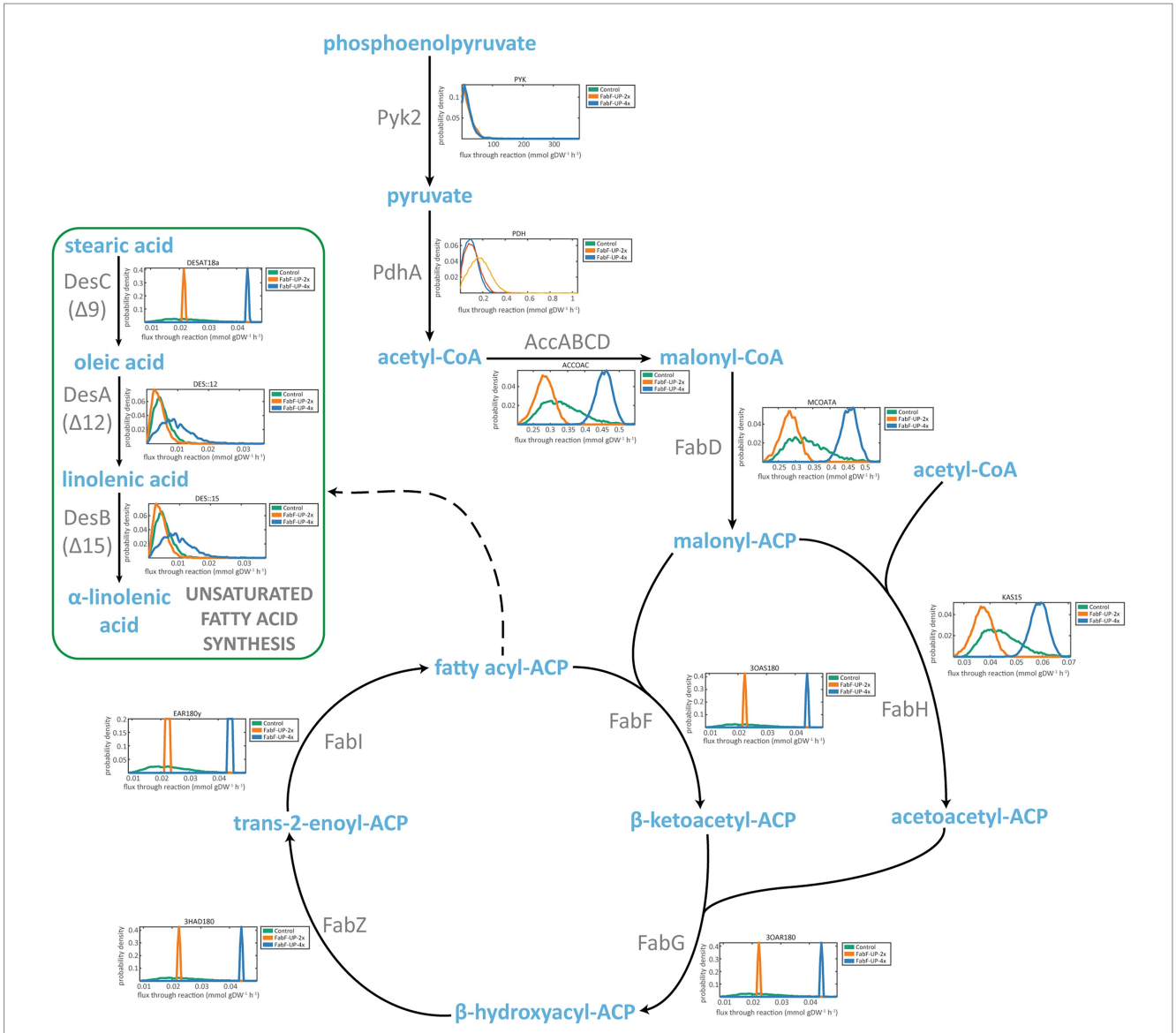

FIGURE 3
Flux sampling distributions of main reactions of ALA synthesis for *in silico* FabF overexpression mutants and predicted using *i*MS837. For each reaction, a plot of the probability density versus the predicted flux of the specific reaction is represented. In green, the results obtained for *i*MS837_ALA are represented (Control); in orange, an *in silico* designed strain that has double flux for FabF reaction (3OAS180) than *i*MS837_ALA (FabF-UP-2x); and in blue, an *in silico* designed strain that has quadruple flux for FabF reaction (3OAS180) than *i*MS837_ALA (FabF-UP-4x).

producing desirable compounds. OptKnock (Burgard et al., 2003) and GDLS (Lun et al., 2009) are strain design algorithms commonly used to predict genetic manipulations for target overproduction. Both methods are based on constraint-based optimization processes to suggest reaction knockout interventions (constraining the metabolic flux of a reaction to zero) to increase targeted compound production while optimizing biomass yield and product yield. Once identified, the suggested reactions can be eliminated *in vivo* by knocking out one or more of the genes encoding the enzymes catalyzing the reaction.

Unfortunately, the application of GDLS and OptKnock algorithms failed identifying suitable knockout strategies for coupling ALA production to biomass synthesis. However, both algorithms were able to identify partial strategies harboring an increasing number of knockouts putatively resulting in ALA overproduction (Figure 6). Overall, and as could be expected, a higher number of knockouts resulted in a higher ALA production due the removal of competing pathways. Interestingly, Optknock was shown as the most efficient algorithm under the condition tested being able to identify more potential knockout target reactions than GDLS (Supplementary Figure S2; Table 2). Among the six reactions identified as potential target to improve ALA production by both algorithms, four of them were related to amino acid metabolism (ALAD_L, PSERT, PGCD, and PSP_L) and two were involved in the oxidative branch of the Pentose Phosphate Pathway (G6PDH2r and GND, respectively. As mentioned in the previous section, the synthesis of these amino acids outcompetes for Acetyl-CoA necessary for fatty acid biosynthesis. A similar explanation could be attributed to the other two common reactions obtained using OptKnock and GDLS, G6PDH2r and GND, since they are diverting glucose away to the formation of Acetyl-CoA. OptKnock was able to identify a

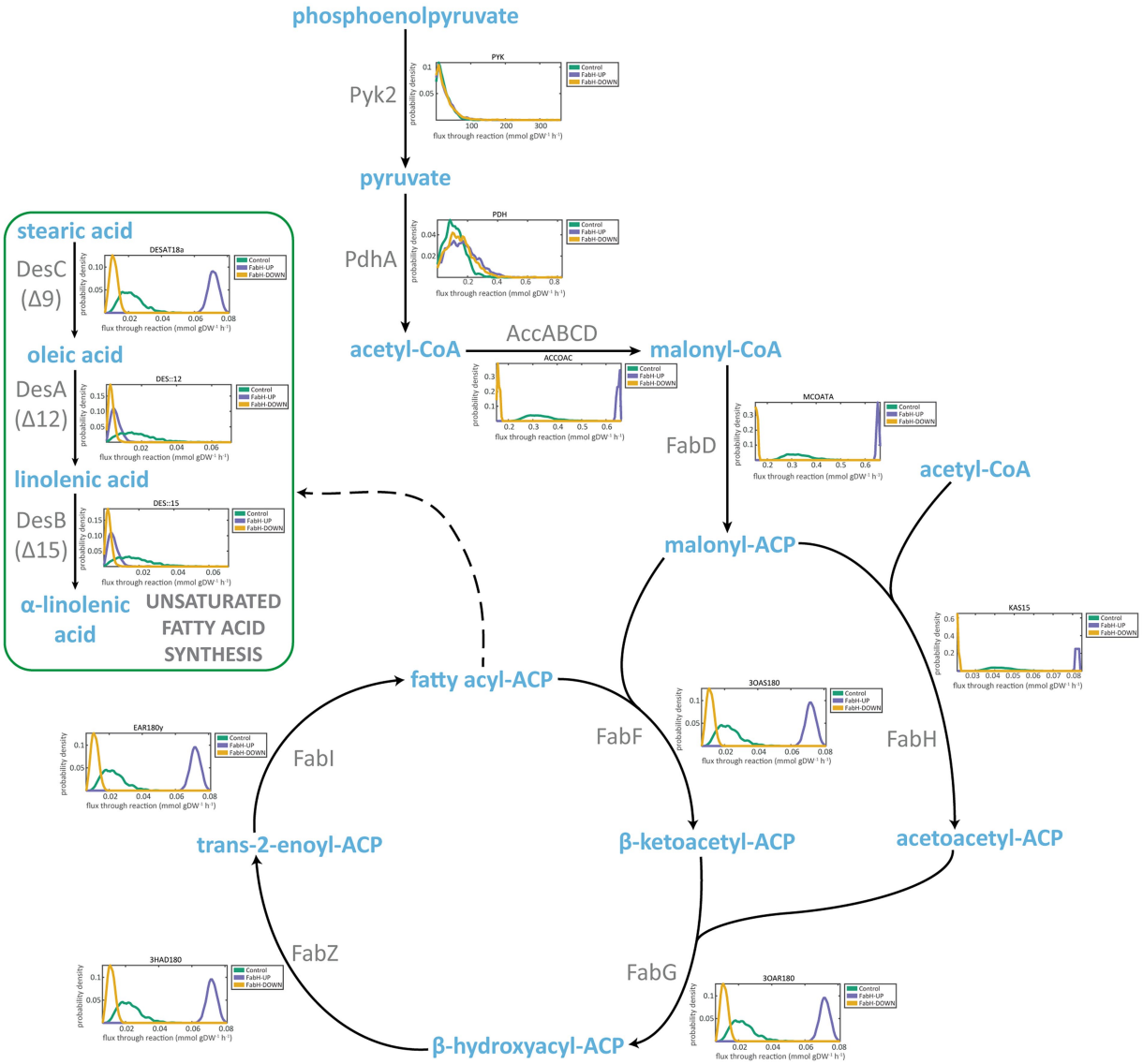

FIGURE 4
Flux sampling distributions of main reactions of ALA synthesis for *in silico* FabH overexpression and downregulation mutants and predicted using *i*MS837. For each reaction, a plot of the probability density versus the predicted flux of the specific reaction is represented. In green, the results obtained for *i*MS837_ALA are represented (Control); in purple, an *in silico* designed strain that has double flux for FabH reaction (KAS15) than *i*MS837_ALA (FabH-UP); and in yellow, an *in silico* designed strain that has half flux for FabH reaction (KAS15) than *i*MS837_ALA (FabH-DOWN).

reaction involved in the fatty acid synthesis as a potential target to be knocked out, ACOATA (Supplementary Figure S2; Table 2). This reaction is performed for FabH, an enzyme that failed *in silico* (Figure 4) and experimentally to increase ALA production (Santos-Merino et al., 2018).

The optimal solutions for most of the single, double, triple, and quadruple knockouts were different for OptKnock and GDLS predictions (Supplementary Table S5; Figures 6A–C). Only in the case of some triple knockouts, the knockouts in the identified reactions gave identical optimal changes in the flux distribution (Supplementary Table S5; Figure 6B). In addition, only the GDLS algorithm was able to suggest optimal solutions by knocking out five reactions (Supplementary Table S5; Figure 6D). Finally, none of the algorithms was able to provide optimal solutions by applying single reaction knockout strategies.

Although growth coupled overproducing strategies are challenging due the low metabolic robustness of cyanobacteria (Nogales et al., 2013; Gudmundsson and Nogales, 2015), which could result in unfeasible genetic designs under the current scenario, we cannot rule out the possibility that the lack of success using GDLS and OptKnock is due to an insufficiently scrutinized metabolic space. To address a more systematic search, we performed a new strain designing analysis by using gcFront (Legon et al., 2022). gcFront is an algorithm that explores knockout strategies maximizing not only cell growth and product synthesis, but also the strength of production-to-growth coupling using a tri-level optimization. The incorporation of this last optimization parameter significantly reduces the search time with respect other strain designing algorithms such as OptKnock or GDLS, thus significantly speeding up the process. In addition, gcFront is based on a genetic algorithm approach, thus, it allows to perform a

larger search in terms of number of knockouts. Unfortunately, we were not able to find growth-coupled ALA overproducing strategies even allowing up to 30 knockouts (data not shown). Taking together, these results confirmed the limited chance of designing growth-coupled ALA overproducing phenotypes using cyanobacteria under photoautotrophic conditions.

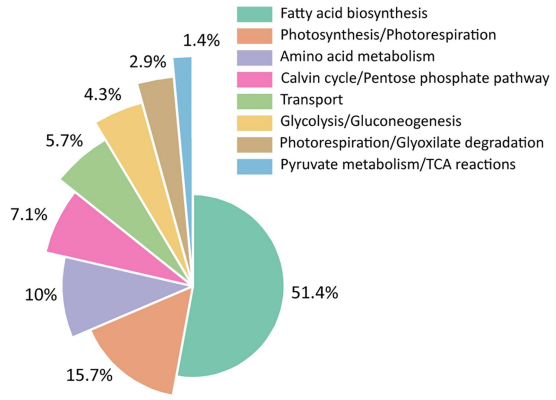

FIGURE 5
Distribution of reactions obtained with FSEOF algorithm based on functional subsystems. The reactions included in each subsystem are depicted in Supplementary Table S4.

2.6. Exploring the production of ALA in *S. elongatus* PCC 7942 beyond photoautotrophic conditions

The lack of growth coupled ALA designs under photoautotrophic conditions encouraging us to explore alternative nutritional regimens, such as photomixotrophy. It has been previously demonstrated that photomixotrophic conditions can enhance growth performance of *S. elongatus* PCC 7942 while provide, at least in theory, a more robust metabolism increasing the metabolic space suitable to flux rerouting (Yan et al., 2012; Nogales et al., 2013). To address this goal, we constructed a set of condition-specific GEMs using as constraint the uptake of inorganic and organic carbon sources previously reported (Yan et al., 2012), as detailed in methods. Following this procedure, we construct a photoautotrophic model and two photomixotrophic GEMs harboring acetate and glucose consumption systems, respectively. The ALA production envelope of the

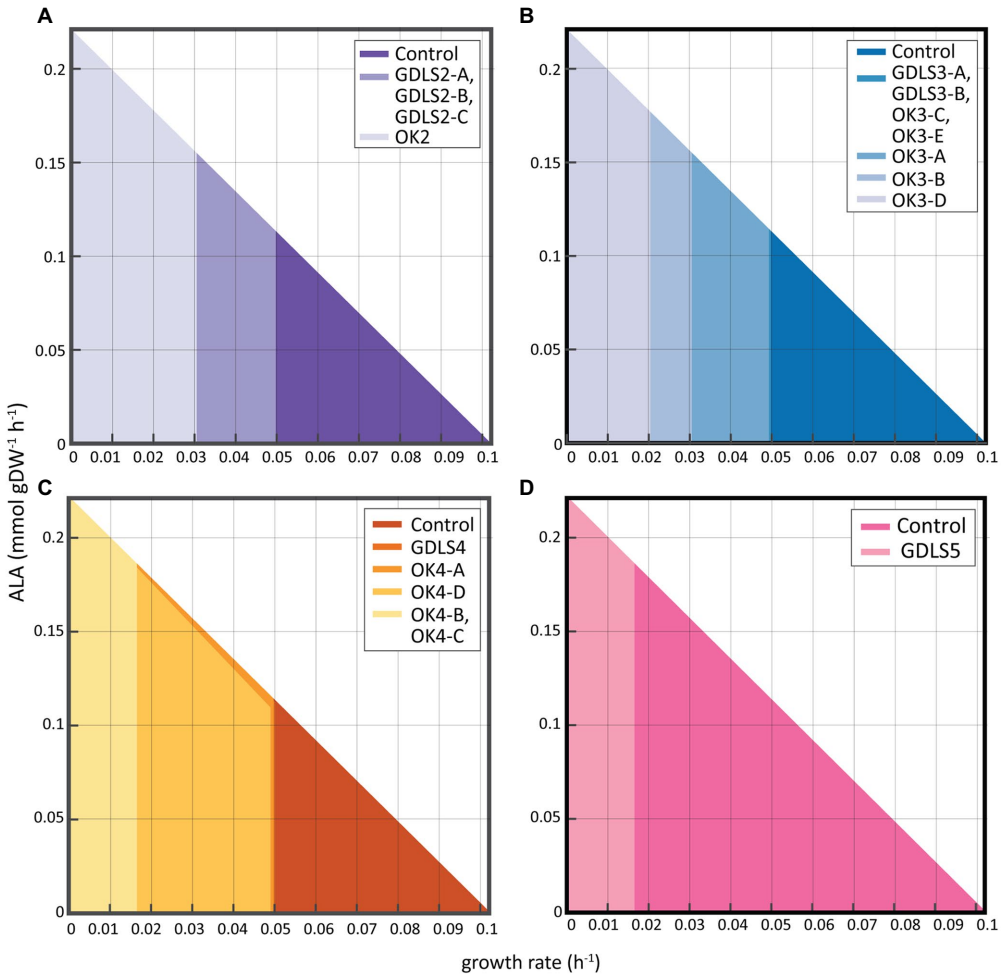

FIGURE 6
Production envelope for *S. elongatus* PCC 7942 mutants with enhanced ALA production obtained with GDLS and OptKnock. Maximum optimal production rate (mmol gDW^{-1} h^{-1}) of ALA achievable with **(A)** two knockouts; **(B)** three knockouts; **(C)** four knockouts; and **(D)** five knockouts.

TABLE 2 Comparison of the reactions suggested by OptKnock and GDLS to be deleted in order to increase the production of ALA.

Method	Suggested reactions	Enzyme name	Pathway
OptKnock	ACOATA	Acetyl-CoA ACP transacylase	Fatty acid biosynthesis
OptKnock	AGDI	Agmatine deiminase	Amino acid metabolism
GDLS/OptKnock	ALAD_L	L-alanine dehydrogenase	Amino acid metabolism
GDLS	ALCD1	Alcohol dehydrogenase (glycerol)	Glycolysis/Gluconeogenesis
OptKnock	FALGTHLs	Formaldehyde glutathione ligase	Cofactor biosynthesis
OptKnock	FBA	Fructose-bisphosphate aldolase	Glycolysis/Gluconeogenesis
OptKnock	FUM	Fumarase	Amino acid metabolism
GDLS/OptKnock	G6PDH2r	Glucose 6-phosphate dehydrogenase	Calvin cycle/Pentose phosphate pathway
OptKnock	GART	GAR transformylase-T	Purine and Pyrimidine Biosynthesis
OptKnock	GHMT2r	Glycine hydroxymethyltransferase	Amino acid metabolism
GDLS/ OptKnock	GND	Phosphogluconate dehydrogenase	Calvin cycle/Pentose phosphate pathway
OptKnock	H2CO3_NAt_syn	Sodium/bicarbonate symporter (SbtA)	Transport: Inner Membrane
OptKnock	PDS2	Phytofluene dehydrogenase	Carotenoid biosynthesis
GDLS/OptKnock	PGCD	Phosphoglycerate dehydrogenase	Amino acid metabolism
GDLS/OptKnock	PGL	6-phosphogluconolactonase	Calvin cycle/Pentose phosphate pathway
GDLS/OptKnock	PSERT	Phosphoserine transaminase	Amino acid metabolism
GDLS	PSP_L	Phosphoserine phosphatase (L-serine)	Amino acid metabolism
OptKnock	PYK	Pyruvate kinase	Pyruvate metabolism/TCA Reactions
GDLS	VALTA	Valine transaminase	Amino acid metabolism

condition-specific models revealed a significant higher metabolic space bounded by photomixotrophic conditions (Figure 7A). Glucose provided the highest metabolic solution space with up to double production of ALA and growth rate, whereas acetate provided a slightly chance of ALA production although it was not growth-coupled. This improved phenotypic performance under photomixotrophic conditions was not only due the presence of organic carbons as additional nutrients, but also to improved photosynthetic efficiencies (Figure 7B). In fact, we computed significant higher fluxes through Photosystem I and II reactions as well as a higher oxygen evolution and photon uptake under photomixotrophic conditions, completely agreeing experimental data (Yan et al., 2012). Subsequently, we used this expanded metabolic space to search for growth-coupled strategies using gcFront following identical setup than used under autotrophic conditions. Despite several attempts performed, we were not able to identify growth-coupled ALA-overproducing *S. elongatus* PCC 7942 strains neither using glucose nor acetate as organic carbon sources (data not shown). Therefore, we concluded that the single use of knockout strategy is not feasible to reroute carbon flux toward the production of ALA.

To gain further insights on this hypothesis, we analyzed the metabolic flux of the reactions from central metabolism including Calvin–Benson–Bassham (CBB) cycle and TCA and those directly involved in ALA synthesis under photoautotrophic and photomixotrophic conditions (Supplementary Figures S3–S5). The detailed analysis of this flux distribution identified that the synthesis of ALA is limited since Acetyl-CoA pool is funneled almost exclusively to the synthesis of fatty acids. In fact, the metabolism of fatty acids *via* Acetyl-CoA is not connected with the production of other components of the biomass. Compounding the problem, the complete list of genes involved in fatty acid biosynthesis are essential under the three nutritional regimens analyzed, excluding the possibility to reroute carbon flux from fatty acid biosynthesis to ALA production *via* removing competitive pathways. Taking together, the reduced connectivity of Acetyl-CoA and the essentiality of genes surrounding fatty acid biosynthesis and ALA production explains, at great extent, the unfeasibility of designing growth coupled ALA overproducer *S. elongatus* PCC 7942 strains.

As a direct consequence, the only solution found increasing ALA production was the increase of the flux through the Δ12-desaturase (DES::12) reaction (Figure 8; Supplementary Figure S3), as it has been demonstrated *in vivo* (Sakamoto et al., 1997; Chen et al., 2014; Santos-Merino et al., 2018). Interestingly, while under photoautotrophic and glucose-driven photomixotrophic conditions the increase in DES::12 flux negatively impacted *S. elongatus* PCC 7942 growth rate (Figure 8; Supplementary Figure S5), we observed an increase in ALA production without any negative effect over the biomass using acetate-driven photomixotrophic conditions (Figure 8; Supplementary Figure S4). This is because, the drainage of the Acetyl-CoA pool toward the ALA production is replenished by increasing the uptake of acetate conditions. Therefore, our computational analysis expanding the metabolic space by feeding *S. elongatus* PCC 7942 with organic carbon sources, strongly suggested that such photomixotrophic conditions, especially using acetate as a carbon source, seem to be a promising strategy to increase ALA production in *S. elongatus* PCC 7942. However, these results will need to be further validated experimentally.

3. Discussion

We generated an updated GEM of *S. elongatus* PCC 792 with considerable improvements in model annotation and accuracy of

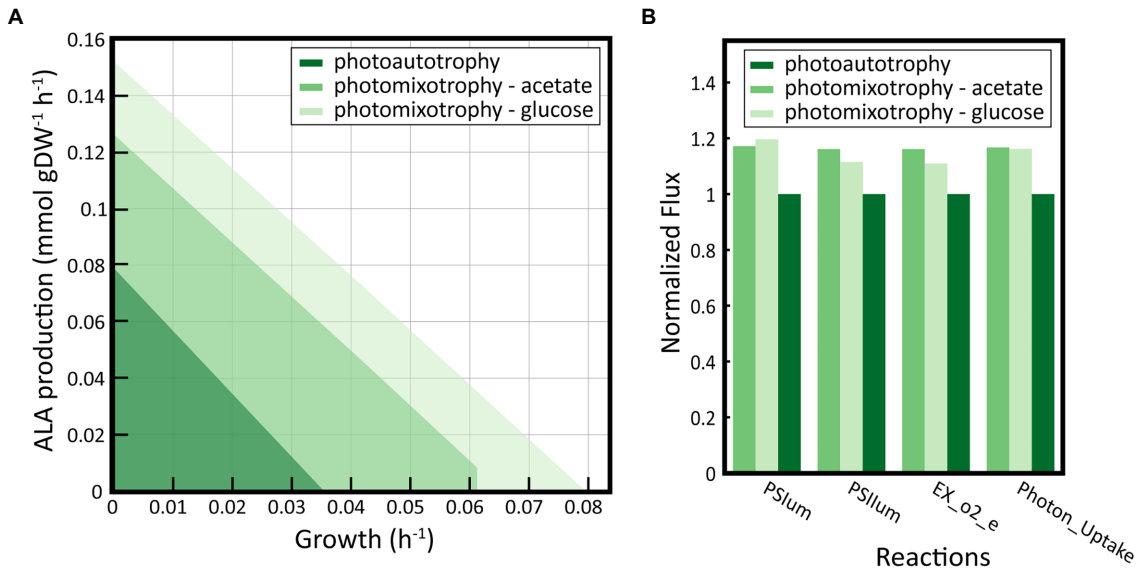

FIGURE 7
Exploration of the metabolic space under photomixotrophic conditions. **(A)** ALA production envelopes in photoautotrophic and photomixotrophic conditions using acetate or glucose as carbon source. The solution space increases under photomixotrophic conditions. **(B)** Evaluation of photosynthetic activity under photoautotrophic and photomixotrophic conditions. The flux of the reactions associated with the activity of Photosystem I and II (PSIum and PSIIum, respectively), the production of oxygen (EX_o2_e) and the flux oh photons (Photon_uptake) under photomixotrophic were simulated and normalized to the values obtained under photoautotrophic conditions.

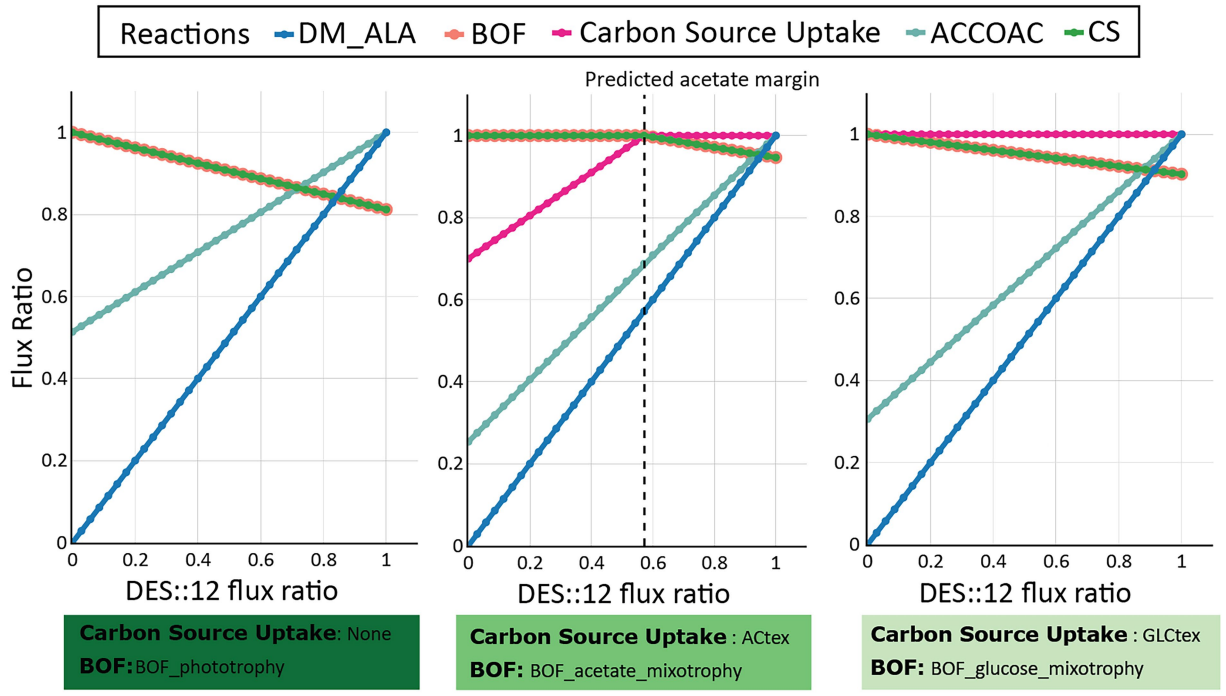

FIGURE 8
Flux variability analysis of Δ12-desaturase (DES::12) overexpression in different nutritional conditions. Impact of DES::12 overexpression on growth rate (BOF), ALA production (DM_ALA), and organic carbon uptake (Carbon Source Uptake) under photoautotrophic (left panel), photomixotrophic with acetate (middle panel), and photomixotrophic with glucose (right panel) conditions. Fluxes through some reactions of the fatty acid synthesis (ACCOAC) and TCA cycle (CS) were measured.

essentiality prediction. The metabolic reaction and genome coverage of the reconstruction was expanded, the format and annotations were updated to be consistent with current best practices. Model improvements were quantified through various metrics such as accuracy of growth yield and ALA yield predictions as well as MEMOTE benchmarking. Overall, *i*MS837 has increased the coverage

of the metabolic functionality of *S. elongatus* PCC 7942 and is one of the highest-quality cyanobacterial GEMs. While the updates included in *i*MS837 largely improved the model accuracy, there were still incorrect predictions about gene essentiality that were not able to be addressed. In *S. elongatus* PCC 7942, a significant proportion of genes has still an unknown function (Labella et al., 2020), with many of them presumably involved in processes relevant to the metabolism of the cyanobacteria. It makes it difficult to include new metabolic genes in the model, since there is a lack of information between GPR associations that will allow their identification. In addition, *i*MS837 only contains a small portion of the total genes of *S. elongatus* PCC 7942 genome (837 of 2,772 genes; Table 1) with an absence of regulation that could influence the correct prediction of their essentiality. However, despite all this missing information, *i*MS837 was able to achieve a high overall accuracy for predicting gene essentiality phenotypes (i.e., 85.5%) that is comparable to the performance of the well-curated *E. coli* models, *i*JO1366 (i.e., 93.4%) (Orth et al., 2011) and *i*ML1515 (i.e., 93.4%) (Monk et al., 2017).

Although *i*MS837 only captures 30.2% of the protein coding genes of *S. elongatus* PCC 7942, this metabolic reconstruction was able to make correct computational predictions related to experimental data (i.e., gene essentiality and fatty acid biosynthesis). The correct functional annotation of the genes encoding metabolic enzymes involved in the fatty acid biosynthetic pathway helped with prediction involving enzymes of this pathway (Supplementary Dataset 1). An experimental effort to identify unknown GPR associations, as well as a constant improvement and updating of the GEM of *S. elongatus* PCC 7942 will improve its prediction capabilities and its use to generate new hypotheses and to identify promising targets for bioengineering applications (Esvelt and Wang, 2013). Integration of kinetics and omics data in GEMs will broaden their quality and application scopes to better understanding the metabolism of cyanobacteria (Gu et al., 2019). On the other hand, the incorporation of accurate and well-developed GEMs into Design-Build-Test-Learn cycles, together with the use of machine learning, will lead to a very powerful toolset for guiding metabolic engineering of cyanobacteria (Liao et al., 2022).

Metabolic models through the implementation of different algorithms are powerful tools to predict potential interventions that may improve the production of a specific compound. FSEOF and GDLS/OptKnock turned out as efficient systems to predict which gene overexpressions and knockouts, respectively, that might be potential targets to increase ALA production. FSEOF results suggest that the availability of saturated fatty acid pools is important for the synthesis of ALA (Figure 5; Supplementary Table S4). This hypothesis was tested previously *in vivo* demonstrated that FabF seems to be the limiting-rate step in this pathway (Santos-Merino et al., 2018). All the solutions obtained with GDLS, OptKnock and gcFront algorithms failed to produce mutants where ALA production is coupled with growth under photoautotrophic conditions (Figure 6).

It has been previously demonstrated that the carbon flux rerouting to obtain growth-coupled producer strains is more challenging under autotrophic conditions than under mixotrophic or heterotrophic conditions in cyanobacteria (Nogales et al., 2013; Wan et al., 2015). Photomixotrophic culture conditions has been successfully applied in *S. elongatus* PCC 7942 to efficiently increase the production of 2,3-butanediol (McEwen et al., 2013; Kanno et al., 2017), which has been also tested *in silico* using *i*JB792 (Broddrick et al., 2019). In addition, the use of acetate as carbon source has allowed to increase the production of poly-3-hydroxybutyrate (PHB) in different cyanobacterial strains (de Philippis et al., 1992; Wu et al., 2002; Sharma and Mallick, 2005; Panda et al., 2006; Khetkorn et al., 2016; Towijit et al., 2018). Here, we successfully demonstrated that the production of ALA can be also boosted *in silico* under photomixotrophic conditions using acetate or glucose as carbon sources (Figure 7A), and this strategy could be employed *in vivo* likely leading to the same result. The synthesis of PHB and ALA requires the same precursor, Acetyl-CoA. The stimulatory effect of acetate on ALA synthesis could be explained by the direct utilization of acetate to increase the intracellular Acetyl-CoA pool, as it has been previously speculated for PHB (de Philippis et al., 1992; Khetkorn et al., 2016).

Overall, we were able to find potential solutions *in silico* that were feasible to increase ALA production in *S. elongatus* PCC 7942 under phototrophic conditions, but in all the cases, they have a negative cost for the cells, negatively impacting their ability to grow. Omega-3 fatty acid synthesis takes place under a continuous supply of Acetyl-CoA and NADPH, which limits carbon flux through the biomass synthesis (Diao et al., 2020). Using *i*MS837, we were able to predict that there is a limitation of the flux of Acetyl-CoA that impedes maximizing ALA production without decreasing biomass yields (Supplementary Figure S3). Contrary to heterotrophic conditions, phototrophic growth promotes low level of Acetyl-CoA since reducing equivalents are provided by photosynthesis instead of TCA *via* oxidation of Acetyl-CoA. Therefore, Acetyl-CoA is mainly used as building block for fatty acids resulting a narrow window for ALA production (Supplementary Figure S3). The use of an organic carbon substrate to stimulate the production of Acetyl-CoA, such as acetate, was the only alternative option to balance the production of ALA and biomass. The utilization of acetate by *S. elongatus* PCC 7942 does not compete with TCA cycle and CO_2 fixation activities (Supplementary Figure S4), but also it seems to stimulate *S. elongatus* PCC 7942 photosynthetic activity (Figure 7B), and potentially the NADPH production, needed for fatty acid production.

4. Conclusion

In this study, we provided and updated GEM of *S. elongatus* PCC 7942 by using information from the scientific literature and openly available databases, as well as data from well-annotated GEMs from other bacteria. The updated model, *i*MS837, comprised 837 genes, 887 reactions, and 801 metabolites. Following a series of growth simulations, the model was found to agree with published literature. The application of the updated GEM to investigate ALA production recapitulated phenotypes observed in literature, and the use of algorithms to identify potential reactions target to be overexpressed or eliminated can offer systematic strategies that would be difficult to delineate experimentally. In addition, photomixotrophic conditions were also identified as potential target to boost ALA production. This application of the updated reconstruction serves as an example of how GEMs can provide insights into non-intuitive metabolic engineering strategies to improve the production of industrially important metabolites. Ultimately, our computational in-depth analysis of *i*MS837 for ALA production provides an example of the systems-biology science iteration paradigm, by producing further hypothesis that need experimental follow-up to be validated.

5. Materials and methods

5.1. Development of an upgraded GEM of *S. elongatus* PCC 7942 named *i*MS837

The *i*MS837 model was developed using the published *i*JB792 model (Broddrick et al., 2019) as a starting point (Supplementary Dataset 3). Updates in *i*JB792 were made using Python and the COBRApy package (Ebrahim et al., 2013). A detailed description of all these changes can be found in the Supplementary Dataset 1 and Supplementary Table S1. The new generated model was denominated *i*MS837. All the scripts used in the methods section can be found in GitHub repository (https://github.com/MariaS87/GEM-Synechococcus-elongatus-PCC-7942-iMS837.git).

5.2. Flux balance analysis

Flux Balanced Analysis (FBA) uses linear programming to maximize an objective function while assuming no metabolite accumulation during cellular growth. We used FBA to evaluate the biomass production (growth prediction) once the biomass reaction was fixed as the objective function (BOF, Biomass Objective Function) (Orth et al., 2010). The result when executing FBA was the growth rate (h^{-1}) predicted under the specified media conditions.

5.3. Model manipulation to produce ALA

*i*MS837 was modified to introduce the required reactions to produce ALA. A detailed list of the reactions and metabolites added is included in Supplementary Table S3. The new generated model was denominated *i*MS837_ALA (Supplementary Dataset 3).

5.4. Metabolic network simulations

*i*MS837 and *i*MS837_ALA models were analyzed using COBRA Toolbox v2.0 (Schellenberger et al., 2011) within the MATLAB environment (The MathWorks Inc.). Tomlab CPLEX (Tomlab Optimization Inc., San Diego, CA) and Gurobi (Gurobi Optimization Inc., Houston, TX) were used for solving the linear programming problems.

5.5. Gene essentiality predictions of *i*MS837

For growth simulation, the biomass equation (BOF) was set as the objective function. The analysis of gene essentiality was performed using the "single_gene_deletion" function of COBRApy (Supplementary Dataset 1; Ebrahim et al., 2013). If the growing rate of the knockout strain was lower than 10^{-3}, the gene was defined as essential.

To evaluate the performance of our GEM to correctly predict gene essentiality, we employed a variety of statistical index-based methods, including: accuracy, sensitivity, specificity, precision, F1-score, and Cohen's Kappa coefficient (van Stralen et al., 2009; Aromolaran et al., 2021). All the statistical metrics were computed based on the scores from true positives (TP), true negatives (TN), false positives (FP), and false negatives (FN). TP and TN occur when both the model prediction and the experimental data agree that a gene is essential and non-essential, respectively. FP occur when the model says a gene is essential, but experiments suggest otherwise, whereas FN occur when the model says a gene is non-essential, but experiments indicate that it is essential (Becker and Palsson, 2008). The aforementioned statistical index-based metrics are described from Equations 1–6 as follows:

$$\%\ \text{accuracy} = \left(\frac{TP + TN}{TP + TN + FP + FN}\right) \cdot 100 \quad (1)$$

$$\%\ \text{sensitivity} = \left(\frac{TP}{TP + FN}\right) \cdot 100 \quad (2)$$

$$\%\ \text{specificity} = \left(\frac{TN}{TN + FP}\right) \cdot 100 \quad (3)$$

$$\%\ \text{precision} = \left(\frac{TP}{TP + FP}\right) \cdot 100 \quad (4)$$

$$\text{F1} - \text{score} = \frac{2 \cdot TP}{2 \cdot TP + FP + FN} \quad (5)$$

$$\text{Kappa} = \frac{2 \cdot (TP \cdot TN - FN \cdot FP)}{\left(\begin{array}{l} TP \cdot FN + TP \cdot FP + 2 \cdot TP \cdot TN + FN \cdot FN \\ + FN \cdot TN + FP \cdot FP + FP \cdot TN \end{array}\right)} \quad (6)$$

The accuracy measures the degree of correctness of a model with respect to both positive and negative classes. The sensitivity estimates the proportion of essential genes that have been correctly identified, whereas the specificity measures the proportion of true negatives that have been correctly predicted. The precision calculates the probability that the essential genes are correctly predicted. The F1-score represents the harmonic mean between precision and sensitivity, combining these two parameters into a single measure (Ghasemian et al., 2022). Lastly, Cohen's Kappa coefficient measures the degree of agreement between the output of experimental versus predicted essentiality data. If Kappa = 1, then the predictions are in perfect agreement with experimental data, and Kappa = 0 means there is no agreement between predictions and experimental data (Aromolaran et al., 2021).

5.6. Monte Carlo flux sampling

The distribution of feasible fluxes in the condition-specific models was calculated by Markov chain Monte Carlo sampling (Schellenberger and Palsson, 2009) implemented in COBRA package (Schellenberger et al., 2011).

5.7. Identification of gene overexpression targets for ALA overproduction

The identification of gene amplification targets was based on the strategy of flux scanning based on enforced objective flux (FSEOF) (Choi et al., 2010). We first simulated the growth behavior of the strain using FBA and set the biomass-producing reaction as the objective function. Then, the maximum theoretical ALA production was obtained by setting the ALA exchange reaction (EX_ALA(e)) as the objective function. In the next steps, this reaction was raised stepwise to reach 80% of the theoretical maximum.

5.8. Identification of potential knockout targets for ALA overproduction

OptKnock (Burgard et al., 2003) and GDLS (Lun et al., 2009) algorithms were implemented to predict potential genetic knockout manipulations that can lead to ALA overproduction. Whereas OptKnock uses bi-level optimization strategies to solve the conflict of cell growth and maximum bioengineering objective, the GDLS algorithm employs reduced metabolic models and predicts gene knockouts based on Gene–Protein-Reaction associations (Xu Z. et al., 2013). Before using these two algorithms, GEM was reduced, including only nonblocked reactions catalyzed by proteins whose genes are nonessential, reactions not involved in transport and reactions with known GPR associations. This step generated a 'reduced' model. These methods were accessible through the COBRA Toolbox v2.0 in MATLAB. For both optimization methods, ALA production flux was set as the optimization target. Each reaction elimination design solution was examined by making the identified changes on bounds to the reactions obtained with OptKnock or GDLS, and were plotted using metabolic production envelope that represents the accessible flux space onto the plane of growth rate versus the target's production rate (Edwards et al., 2002). The suggested reactions obtained with these algorithms can be removed *in vivo* by knocking out one or more of the genes encoding the enzymes catalyzing the reaction.

In addition, the recently developed gcFront algorithm was also used to identify knockouts that growth-couple synthesis (Legon et al., 2022). gcFront uses a multiobjective genetic algorithm that identifies a Pareto front of designs that maximize growth rate, product synthesis and coupling strength and finds combinations of gene/reaction knockouts that will enforce growth coupling (Legon et al., 2022). Before applying this algorithm, GEMs need to be pre-processed to reduce the search space of reaction by removing the biomass reactions not assigned as objective and blocked reactions. All blocked reactions were identified using flux variability analysis (FVA) (Mahadevan and Schilling, 2003) as reactions unable to carry flux when the biomass was constraint to 20%. To reduce computation time by gcFront algorithm, the list of reactions previously identified with Optknock screening with 1–5 maximum number of knockouts was used. Then, gcFront was executed for each nutritional condition (i.e., phototrophic, photomixotrophic with acetate and photomixotrophic with glucose) using the parameter setup described in Table 3.

TABLE 3 List of parameters used with gcFront algorithm.

Parameters	Values for		
	*i*MS837	*i*MS837_acetate	*i*MS837_glucose
biomassrxn	BOF_photoautotrophy	BOF_acetate_mixotrophy	BOF_glucose_mixotrophy
mingrowth	0.003	0.006	0.008
skipreduction	True	True	True
maxreductionsize	30	30	30
ignorelistrxns	R-C	R-C	R-C
popsize	1,000	1,000	1,000
genlimit	100,000	100,000	100,000

5.9. Generation of specific GEMs for photomixotrophic conditions from *i*MS837_ALA

Reactions for the uptake of glucose and acetate were added to *i*MS837_ALA. The experimental data obtained under photomixotrophic conditions with acetate and glucose *in vivo* (Yan et al., 2012) were used to introduce constraints to *i*MS837_ALA to simulate these conditions *in silico*. A summary of these constraints is depicted in Table 4. In addition, a biomass equation was specifically generated for photomixotrophic conditions with acetate and glucose, denominated BOF_acetate_photomixotrophy and BOF_glucose_photomixotrophy, respectively. In order to do that, a previously published protocol to generate biomass objective functions from experimental data was used (Lachance et al., 2019). Condition-specific macromolecular data related to protein, lipid, carbohydrate, and photosynthetic pigment composition were obtained for previously published data (Yan et al., 2012). In addition, transcriptomic data and lipid profiles were also needed and obtained from different literature sources (Rós et al., 2013; Xiong et al., 2015).

5.10. Generation of pathway maps for visualization of metabolic fluxes

Escher was used for visualizing the fluxes of the metabolic pathways involved in ALA synthesis (King et al., 2015). The Escher website was used to draw all the metabolic maps.

Data availability statement

The original contributions presented in the study are included in the article/Supplementary material, further inquiries can be directed to the corresponding authors.

Author contributions

MS-M and JN conceived and designed the study. MS-M and JN prepared the first draft of the manuscript. MS-M performed the metabolic network reconstruction. MS-M and AG-B performed

TABLE 4 Values of the lower and upper bounds used to generate condition-specific GEMs.

Reactions	Bound values for		
	Photoautotrophy	Acetate photomixotrophy	Glucose photomixotrophy
RBPCcx	(1.46, 1,000)	(1.8, 1,000)	(1.73, 1,000)
ACtex	(0, 1,000)	(0.257, 1,000)	(0, 1,000)
GLCtex	(0, 1,000)	(0, 1,000)	(0.175, 1,000)

model simulations. All the authors discussed the results and participated in the writing process.

Funding

This work was supported by the European Union's Horizon 2020 Research and Innovation Programme under Grant agreement no. 101000733 (Promicon) by the Spanish Ministry of Science and Innovation (MICINN) grants RobExplode PID2019-108458RB-I00 (AEI/10.13039/501100011033) and TED2021-130689B-C33 to JN, and PID2020-117923GB-I00 to FdlC. Funding was likewise provided by CSIC's Interdisciplinary Platform for Sustainable Plastics toward a Circular Economy+ (PTI-SusPlast+). MS-M was the recipient of a Ph.D. fellowship (BES-2012-057387) from Spanish Ministry of Economy and Competitiveness (MINECO).

Acknowledgments

The authors wish to thank David San León Granado for his help with Figure 2.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Supplementary material

The Supplementary material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fmicb.2023.1126030/full#supplementary-material

References

Aromolaran, O., Aromolaran, D., Isewon, I., and Oyelade, J. (2021). Machine learning approach to gene essentiality prediction: a review. *Brief. Bioinform.* 22:bbab128. doi: 10.1093/bib/bbab128

Becker, S. A., and Palsson, B. O. (2008). Three factors underlying incorrect in silico predictions of essential metabolic genes. *BMC Syst. Biol.* 2:14. doi: 10.1186/1752-0509-2-14

Berla, B. M., Saha, R., Immethun, C. M., Maranas, C. D., Moon, T. S., and Pakrasi, H. B. (2013). Synthetic biology of cyanobacteria: unique challenges and opportunities. *Front. Microbiol.* 4:246. doi: 10.3389/fmicb.2013.00246

Broddrick, J. T., Rubin, B. E., Welkie, D. G., Du, N., Mih, N., Diamond, S., et al. (2016). Unique attributes of cyanobacterial metabolism revealed by improved genome-scale metabolic modeling and essential gene analysis. *Proc. Natl. Acad. Sci. U. S. A.* 113, E8344–E8353. doi: 10.1073/pnas.1613446113

Broddrick, J. T., Welkie, D. G., Jallet, D., Golden, S. S., Peers, G., and Palsson, B. O. (2019). Predicting the metabolic capabilities of *Synechococcus elongatus* PCC 7942 adapted to different light regimes. *Metab. Eng.* 52, 42–56. doi: 10.1016/j.ymben.2018.11.001

Burgard, A. P., Pharkya, P., and Maranas, C. D. (2003). Optknock: a bilevel programming framework for identifying gene knockout strategies for microbial strain optimization. *Biotechnol. Bioeng.* 84, 647–657. doi: 10.1002/bit.10803

Chen, G., Qu, S., Wang, Q., Bian, F., Peng, Z., Zhang, Y., et al. (2014). Transgenic expression of delta-6 and delta-15 fatty acid desaturases enhances omega-3 polyunsaturated fatty acid accumulation in Synechocystis sp. PCC6803. *Biotechnol. Biofuels* 7:32. doi: 10.1186/1754-6834-7-32

Choi, H. S., Lee, S. Y., Kim, T. Y., and Woo, H. M. (2010). *In silico* identification of gene amplification targets for improvement of lycopene production. *Appl. Environ. Microbiol.* 76, 3097–3105. doi: 10.1128/AEM.00115-10

De Philippis, R., Sili, C., and Vincenzini, M. (1992). Glycogen and poly-β-hydroxybutyrate synthesis in *Spirulina maxima*. *Microbiology* 138, 1623–1628.

Diao, J., Song, X., Guo, T., Wang, F., Chen, L., and Zhang, W. (2020). Cellular engineering strategies toward sustainable omega-3 long chain polyunsaturated fatty acids production: state of the art and perspectives. *Biotechnol. Adv.* 40:107497. doi: 10.1016/j.biotechadv.2019.107497

Dong, X., He, Q., Peng, Z., Yu, J., Bian, F., Li, Y., et al. (2016). Production of Î³-linolenic acid and stearidonic acid by Synechococcus sp. PCC7002 containing cyanobacterial fatty acid desaturase genes. *Chin. J. Oceanol. Limnol.* 34, 772–780. doi: 10.1007/s00343-016-4369-x

Ebrahim, A., Lerman, J. A., Palsson, B. O., and Hyduke, D. R. (2013). COBRApy: COnstraints-based reconstruction and analysis for python. *BMC Syst. Biol.* 7:74. doi: 10.1186/1752-0509-7-74

Edwards, J. S., Ramakrishna, R., and Palsson, B. O. (2002). Characterizing the metabolic phenotype: a phenotype phase plane analysis. *Biotechnol. Bioeng.* 77, 27–36. doi: 10.1002/bit.10047

Esvelt, K. M., and Wang, H. H. (2013). Genome-scale engineering for systems and synthetic biology. *Mol. Syst. Biol.* 9:641. doi: 10.1038/msb.2012.66

Eungrasamee, K., Miao, R., Incharoensakdi, A., Lindblad, P., and Jantaro, S. (2019). Improved lipid production via fatty acid biosynthesis and free fatty acid recycling in engineered *Synechocystis* sp. PCC 6803. *Biotechnol. Biofuels* 12:8. doi: 10.1186/s13068-018-1349-8

Fekrat, F., Nami, B., Hejazi, M. A., Ghaffari, M. R., and Shahbazi, M. (2021). Correlation network analysis of metabolites reveals the role of nitrogen-containing metabolic stressors in stimulating high-value compounds biosynthesis in Arthrospira platensis. *J. Appl. Phycol.* 9, 1967–1982. doi: 10.1007/s10811-022-02753-8

Galán, B., Santos-Merino, M., Nogales, J., De La Cruz, F., and García, J. L. (2019). "Microbial oils as nutraceuticals and animal feeds" in *Health Consequences of Microbial Interactions with Hydrocarbons, Oils, and Lipids*. ed. H. Goldfine (Cham: Springer International Publishing), 1–45.

Ghasemian, B., Shahabi, H., Shirzadi, A., Al-Ansari, N., Jaafari, A., Geertsema, M., et al. (2022). Application of a novel hybrid machine learning algorithm in shallow landslide susceptibility mapping in a mountainous area. *Front. Environ. Sci.* 10:897254. doi: 10.3389/fenvs.2022.897254

Gu, C., Kim, G. B., Kim, W. J., Kim, H. U., and Lee, S. Y. (2019). Current status and applications of genome-scale metabolic models. *Genome Biol.* 20:121. doi: 10.1186/s13059-019-1730-3

Gudmundsson, S., and Nogales, J. (2015). Cyanobacteria as photosynthetic biocatalysts: a systems biology perspective. *Mol. BioSyst.* 11, 60–70. doi: 10.1039/C4MB00335G

Gudmundsson, S., and Nogales, J. (2021). Recent advances in model-assisted metabolic engineering. *Curr. Opin. Syst. Biol.* 28:100392. doi: 10.1016/j.coisb.2021.100392

Hendry, J. I., Bandyopadhyay, A., Srinivasan, S., Pakrasi, H. B., and Maranas, C. D. (2020). Metabolic model guided strain design of cyanobacteria. *Curr. Opin. Biotechnol.* 64, 17–23. doi: 10.1016/j.copbio.2019.08.011

Herrmann, H. A., Dyson, B. C., Vass, L., Johnson, G. N., and Schwartz, J. M. (2019). Flux sampling is a powerful tool to study metabolism under changing environmental conditions. *NPJ Syst. Biol. Appl.* 5:32. doi: 10.1038/s41540-019-0109-0

Kanno, M., Carroll, A. L., and Atsumi, S. (2017). Global metabolic rewiring for improved CO_2 fixation and chemical production in cyanobacteria. *Nat. Commun.* 8:14724. doi: 10.1038/ncomms14724

Khetkorn, W., Incharoensakdi, A., Lindblad, P., and Jantaro, S. (2016). Enhancement of poly-3-hydroxybutyrate production in *Synechocystis* sp. PCC 6803 by overexpression of its native biosynthetic genes. *Bioresour. Technol.* 214, 761–768. doi: 10.1016/j.biortech.2016.05.014

King, Z. A., Drager, A., Ebrahim, A., Sonnenschein, N., Lewis, N. E., and Palsson, B. O. (2015). Escher: a web application for building, sharing, and embedding data-rich visualizations of biological pathways. *PLoS Comput. Biol.* 11:e1004321. doi: 10.1371/journal.pcbi.1004321

Labella, J. I., Llop, A., and Contreras, A. (2020). The default cyanobacterial linked genome: an interactive platform based on cyanobacterial linkage networks to assist functional genomics. *FEBS Lett.* 594, 1661–1674. doi: 10.1002/1873-3468.13775

Lachance, J. C., Lloyd, C. J., Monk, J. M., Yang, L., Sastry, A. V., Seif, Y., et al. (2019). BOFdat: generating biomass objective functions for genome-scale metabolic models from experimental data. *PLoS Comput. Biol.* 15:e1006971. doi: 10.1371/journal.pcbi.1006971

Legon, L., Corre, C., Bates, D. G., and Mannan, A. A. (2022). gcFront: a tool for determining a Pareto front of growth-coupled cell factory designs. *Bioinformatics* 38, 3657–3659. doi: 10.1093/bioinformatics/btac376

Lewis, N. E., Nagarajan, H., and Palsson, B. O. (2012). Constraining the metabolic genotype-phenotype relationship using a phylogeny of *in silico* methods. *Nat. Rev. Microbiol.* 10, 291–305. doi: 10.1038/nrmicro2737

Liao, X., Ma, H., and Tang, Y. J. (2022). Artificial intelligence: a solution to involution of design-build-test-learn cycle. *Curr. Opin. Biotechnol.* 75:102712. doi: 10.1016/j.copbio.2022.102712

Lieven, C., Beber, M. E., Olivier, B. G., Bergmann, F. T., Ataman, M., Babaei, P., et al. (2020). MEMOTE for standardized genome-scale metabolic model testing. *Nat. Biotechnol.* 38, 272–276. doi: 10.1038/s41587-020-0446-y

Ludwig, M., and Bryant, D. A. (2012). *Synechococcus* sp. strain PCC 7002 Ttranscriptome: acclimation to temperature, salinity, oxidative stress, and mixotrophic growth conditions. *Front. Microbiol.* 3:354. doi: 10.3389/fmicb.2012.00354

Lun, D. S., Rockwell, G., Guido, N. J., Baym, M., Kelner, J. A., Berger, B., et al. (2009). Large-scale identification of genetic design strategies using local search. *Mol. Syst. Biol.* 5:296. doi: 10.1038/msb.2009.57

Mahadevan, R., and Schilling, C. H. (2003). The effects of alternate optimal solutions in constraint-based genome-scale metabolic models. *Metab. Eng.* 5, 264–276. doi: 10.1016/j.ymben.2003.09.002

Mcewen, J. T., Machado, I. M., Connor, M. R., and Atsumi, S. (2013). Engineering *Synechococcus elongatus* PCC 7942 for continuous growth under diurnal conditions. *Appl. Environ. Microbiol.* 79, 1668–1675. doi: 10.1128/AEM.03326-12

Mills, L. A., Mccormick, A. J., and Lea-Smith, D. J. (2020). Current knowledge and recent advances in understanding metabolism of the model cyanobacterium *Synechocystis* sp PCC 6803. *Biosci. Rep.* 40:BSR20193325.

Monk, J. M., Lloyd, C. J., Brunk, E., Mih, N., Sastry, A., King, Z., et al. (2017). *i*ML1515, a knowledgebase that computes *Escherichia coli* traits. *Nat. Biotechnol.* 35, 904–908. doi: 10.1038/nbt.3956

Nogales, J., Gudmundsson, S., Knight, E. M., Palsson, B. O., and Thiele, I. (2012). Detailing the optimality of photosynthesis in cyanobacteria through systems biology analysis. *Proc. Natl. Acad. Sci. U. S. A.* 109, 2678–2683. doi: 10.1073/pnas.1117907109

Nogales, J., Gudmundsson, S., and Thiele, I. (2013). Toward systems metabolic engineering in cyanobacteria: opportunities and bottlenecks. *Bioengineered* 4, 158–163. doi: 10.4161/bioe.22792

Nogales, J., Mueller, J., Gudmundsson, S., Canalejo, F. J., Duque, E., Monk, J., et al. (2020). High-quality genome-scale metabolic modelling of pseudomonas putida highlights its broad metabolic capabilities. *Environ. Microbiol.* 22, 255–269. doi: 10.1111/1462-2920.14843

Nozzi, N. E., and Atsumi, S. (2015). Genome engineering of the 2,3-butanediol biosynthetic pathway for tight regulation in cyanobacteria. *ACS Synth. Biol.* 4, 1197–1204. doi: 10.1021/acssynbio.5b00057

Oberhardt, M. A., Palsson, B. O., and Papin, J. A. (2009). Applications of genome-scale metabolic reconstructions. *Mol. Syst. Biol.* 5:320. doi: 10.1038/msb.2009.77

Oliver, J. W., Machado, I. M., Yoneda, H., and Atsumi, S. (2013). Cyanobacterial conversion of carbon dioxide to 2,3-butanediol. *Proc. Natl. Acad. Sci. U. S. A.* 110, 1249–1254. doi: 10.1073/pnas.1213024110

Orth, J. D., Conrad, T. M., Na, J., Lerman, J. A., Nam, H., Feist, A. M., et al. (2011). A comprehensive genome-scale reconstruction of Escherichia coli metabolism--2011. *Mol. Syst. Biol.* 7:535. doi: 10.1038/msb.2011.65

Orth, J. D., Thiele, I., and Palsson, B. O. (2010). What is flux balance analysis? *Nat. Biotechnol.* 28, 245–248. doi: 10.1038/nbt.1614

Panda, B., Jain, P., Sharma, L., and Mallick, N. (2006). Optimization of cultural and nutritional conditions for accumulation of poly-beta-hydroxybutyrate in *Synechocystis* sp. PCC 6803. *Bioresour. Technol.* 97, 1296–1301. doi: 10.1016/j.biortech.2005.05.013

Park, J. M., Park, H. M., Kim, W. J., Kim, H. U., Kim, T. Y., and Lee, S. Y. (2012). Flux variability scanning based on enforced objective flux for identifying gene amplification targets. *BMC Syst. Biol.* 6:106. doi: 10.1186/1752-0509-6-106

Patel, A., Karageorgou, D., Rova, E., Katapodis, P., Rova, U., Christakopoulos, P., et al. (2020). An overview of potential oleaginous microorganisms and their role in biodiesel and omega-3 fatty acid-based industries. *Microorganisms* 8:434. doi: 10.3390/microorganisms8030434

Poole, L. B., Parsonage, D., Sergeant, S., Miller, L. R., Lee, J., Furdui, C. M., et al. (2020). Acyl-lipid desaturases and Vipp1 cooperate in cyanobacteria to produce novel omega-3 PUFA-containing glycolipids. *Biotechnol. Biofuels* 13:83. doi: 10.1186/s13068-020-01719-7

Rajneesh, , Singh, S. P., Pathak, J., and Sinha, R. P. (2017). Cyanobacterial factories for the production of green energy and value-added products: an integrated approach for economic viability. *Renew. Sust. Energ. Rev.* 69, 578–595. doi: 10.1016/j.rser.2016.11.110

Rós, P., Silva, C. S., Silva-Stenico, M. E., Fiore, M. F., and De Castro, H. F. (2013). Assessment of chemical and physico-chemical properties of cyanobacterial lipids for biodiesel production. *Mar. Drugs* 11, 2365–2381. doi: 10.3390/md11072365

Rubin, B. E., Wetmore, K. M., Price, M. N., Diamond, S., Shultzaberger, R. K., Lowe, L. C., et al. (2015). The essential gene set of a photosynthetic organism. *Proc. Natl. Acad. Sci. U. S. A.* 112, E6634–E6643. doi: 10.1073/pnas.1519220112

Sakamoto, T., Higashi, S., Wada, H., Murata, N., and Bryant, D. A. (1997). Low-temperature-induced desaturation of fatty acids and expression of desaturase genes in the cyanobacterium *Synechococcus* sp. PCC 7002. *FEMS Microbiol. Lett.* 152, 313–320. doi: 10.1111/j.1574-6968.1997.tb10445.x

Santos-Merino, M., Garcillan-Barcia, M. P., and De La Cruz, F. (2018). Engineering the fatty acid synthesis pathway in *Synechococcus elongatus* PCC 7942 improves omega-3 fatty acid production. *Biotechnol. Biofuels* 11:239. doi: 10.1186/s13068-018-1243-4

Santos-Merino, M., Gutiérrez-Lanza, R., Nogales, J., Garcia, J. L., and De La Cruz, F. (2022). *Synechococcus elongatus* PCC 7942 as a platform for bioproduction of omega-3 fatty acids. *Life* 12:810. doi: 10.3390/life12060810

Santos-Merino, M., Singh, A. K., and Ducat, D. C. (2019). New applications of synthetic biology tools for cyanobacterial metabolic engineering. *Front. Bioeng. Biotechnol.* 7:33. doi: 10.3389/fbioe.2019.00033

Schellenberger, J., and Palsson, B. O. (2009). Use of randomized sampling for analysis of metabolic networks. *J. Biol. Chem.* 284, 5457–5461. doi: 10.1074/jbc.R800048200

Schellenberger, J., Que, R., Fleming, R. M., Thiele, I., Orth, J. D., Feist, A. M., et al. (2011). Quantitative prediction of cellular metabolism with constraint-based models: the COBRA toolbox v2.0. *Nat. Protoc.* 6, 1290–1307. doi: 10.1038/nprot.2011.308

Sharma, L., and Mallick, N. (2005). Enhancement of poly-beta-hydroxybutyrate accumulation in *Nostoc muscorum* under mixotrophy, chemoheterotrophy and limitations of gas-exchange. *Biotechnol. Lett.* 27, 59–62. doi: 10.1007/s10529-004-6586-1

Shinde, S., Singapuri, S., Jiang, Z., Long, B., Wilcox, D., Klatt, C., et al. (2022). Thermodynamics contributes to high limonene productivity in cyanobacteria. *Metab. Eng. Commun.* 14:e00193. doi: 10.1016/j.mec.2022.e00193

Tan, X., Yao, L., Gao, Q., Wang, W., Qi, F., and Lu, X. (2011). Photosynthesis driven conversion of carbon dioxide to fatty alcohols and hydrocarbons in cyanobacteria. *Metab. Eng.* 13, 169–176. doi: 10.1016/j.ymben.2011.01.001

Tasaka, Y., Gombos, Z., Nishiyama, Y., Mohanty, P., Ohba, T., Ohki, K., et al. (1996). Targeted mutagenesis of acyl-lipid desaturases in *Synechocystis*: evidence for the important roles of polyunsaturated membrane lipids in growth, respiration and photosynthesis. *EMBO J.* 15, 6416–6425. doi: 10.1002/j.1460-2075.1996.tb01033.x

Towijit, U., Songruk, N., Lindblad, P., Incharoensakdi, A., and Jantaro, S. (2018). Co-overexpression of native phospholipid-biosynthetic genes *plsX* and *plsC* enhances lipid production in *Synechocystis* sp. PCC 6803. *Sci. Rep.* 8:13510.

Triana, J., Montagud, A., Siurana, M., Fuente, D., Urchueguia, A., Gamermann, D., et al. (2014). Generation and evaluation of a genome-scale metabolic network model of *Synechococcus elongatus* PCC7942. *Meta* 4, 680–698. doi: 10.3390/metabo4030680

Van Stralen, K. J., Stel, V. S., Reitsma, J. B., Dekker, F. W., Zoccali, C., and Jager, K. J. (2009). Diagnostic methods I: sensitivity, specificity, and other measures of accuracy. *Kidney Int.* 75, 1257–1263. doi: 10.1038/ki.2009.92

Vavitsas, K., Kugler, A., Satta, A., Hatzinikolaou, D. G., Lindblad, P., Fewer, D. P., et al. (2021). Doing synthetic biology with photosynthetic microorganisms. *Physiol. Plant.* 173, 624–638. doi: 10.1111/ppl.13455

Wan, N., Abernathy, M., Tang, J. K.-H., Tang, Y. J., and You, L. (2015). Cyanobacterial photo-driven mixotrophic metabolism and its advantages for biosynthesis. *Front. Chem. Sci. Eng.* 9, 308–316. doi: 10.1007/s11705-015-1521-7

Wang, W., Liu, X., and Lu, X. (2013). Engineering cyanobacteria to improve photosynthetic production of alka(e)nes. *Biotechnol. Biofuels* 6:69. doi: 10.1186/1754-6834-6-69

Wei, W., Ning, L. W., Ye, Y. N., and Guo, F. B. (2013). Geptop: a gene essentiality prediction tool for sequenced bacterial genomes based on orthology and phylogeny. *PLoS One* 8:e72343. doi: 10.1371/journal.pone.0072343

Wu, G. F., Shen, Z. Y., and Wu, Q. Y. (2002). Modification of carbon partitioning to enhance PHB production in *Synechocystis* sp. PCC6803. *Enzym. Microb. Technol.* 30, 710–715. doi: 10.1016/S0141-0229(02)00044-3

Xiong, Q., Feng, J., Li, S. T., Zhang, G. Y., Qiao, Z. X., Chen, Z., et al. (2015). Integrated transcriptomic and proteomic analysis of the global response of *Synechococcus* to high light stress. *Mol. Cell. Proteomics* 14, 1038–1053. doi: 10.1074/mcp.M114.046003

Xu, C., Liu, L., Zhang, Z., Jin, D., Qiu, J., and Chen, M. (2013). Genome-scale metabolic model in guiding metabolic engineering of microbial improvement. *Appl. Microbiol. Biotechnol.* 97, 519–539. doi: 10.1007/s00253-012-4543-9

Xu, Z., Zheng, P., Sun, J., and Ma, Y. (2013). ReacKnock: identifying reaction deletion strategies for microbial strain optimization based on genome-scale metabolic network. *PLoS One* 8:e72150. doi: 10.1371/journal.pone.0072150

Yan, R., Zhu, D., Zhang, Z., Zeng, Q., and Chu, J. (2012). Carbon metabolism and energy conversion of *Synechococcus* sp. PCC 7942 under mixotrophic conditions: comparison with photoautotrophic condition. *J. Appl. Phycol.* 24, 657–668. doi: 10.1007/s10811-011-9683-2

Yoshino, T., Kakunaka, N., Liang, Y., Ito, Y., Maeda, Y., Nomaguchi, T., et al. (2017). Production of omega3 fatty acids in marine cyanobacterium *Synechococcus* sp. strain NKBG 15041c via genetic engineering. *Appl. Microbiol. Biotechnol.* 101, 6899–6905. doi: 10.1007/s00253-017-8407-1

Zahra, Z., Choo, D. H., Lee, H., and Parveen, A. (2020). Cyanobacteria: review of current potentials and applications. *Environments* 7:13. doi: 10.3390/environments7020013

Zhang, Y. M., and Rock, C. O. (2009). Transcriptional regulation in bacterial membrane lipid synthesis. *J. Lipid Res.* 50, S115–S119. doi: 10.1194/jlr.R800046-JLR200

frontiers | Frontiers in Microbiology

TYPE Original Research
PUBLISHED 22 March 2023
DOI 10.3389/fmicb.2023.1122078

OPEN ACCESS

EDITED BY
Xuefeng Lu,
Qingdao Institute of Bioenergy and Bioprocess Technology (CAS),
China

REVIEWED BY
Peter Lindblad,
Uppsala University,
Sweden
Annegret Wilde,
University of Freiburg,
Germany

*CORRESPONDENCE
Stephan Klähn
✉ stephan.klaehn@ufz.de

SPECIALTY SECTION
This article was submitted to Microbiotechnology, a section of the journal Frontiers in Microbiology

RECEIVED 12 December 2022
ACCEPTED 22 February 2023
PUBLISHED 22 March 2023

CITATION
Opel F, Itzenhäuser MA, Wehner I, Lupacchini S, Lauterbach L, Lenz O and Klähn S (2023) Toward a synthetic hydrogen sensor in cyanobacteria: Functional production of an oxygen-tolerant regulatory hydrogenase in *Synechocystis* sp. PCC 6803. *Front. Microbiol.* 14:1122078. doi: 10.3389/fmicb.2023.1122078

Toward a synthetic hydrogen sensor in cyanobacteria: Functional production of an oxygen-tolerant regulatory hydrogenase in *Synechocystis* sp. PCC 6803

Franz Opel[1], Marvin Amadeus Itzenhäuser[1], Isabel Wehner[1], Sara Lupacchini[1], Lars Lauterbach[2], Oliver Lenz[3] and Stephan Klähn[1]*

[1]Department of Solar Materials, Helmholtz Centre for Environmental Research – UFZ, Leipzig, Germany, [2]Institute of Applied Microbiology (iAMB), RWTH Aachen University, Aachen, Germany, [3]Institute of Chemistry, Technical University of Berlin, Berlin, Germany

Cyanobacteria have raised great interest in biotechnology, e.g., for the sustainable production of molecular hydrogen (H_2) using electrons from water oxidation. However, this is hampered by various constraints. For example, H_2-producing enzymes compete with primary metabolism for electrons and are usually inhibited by molecular oxygen (O_2). In addition, there are a number of other constraints, some of which are unknown, requiring unbiased screening and systematic engineering approaches to improve the H_2 yield. Here, we introduced the regulatory [NiFe]-hydrogenase (RH) of *Cupriavidus necator* (formerly *Ralstonia eutropha*) H16 into the cyanobacterial model strain *Synechocystis* sp. PCC 6803. In its natural host, the RH serves as a molecular H_2 sensor initiating a signal cascade to express hydrogenase-related genes when no additional energy source other than H_2 is available. Unlike most hydrogenases, the *C. necator* enzymes are O_2-tolerant, allowing their efficient utilization in an oxygenic phototroph. Similar to *C. necator*, the RH produced in *Synechocystis* showed distinct H_2 oxidation activity, confirming that it can be properly matured and assembled under photoautotrophic, i.e., oxygen-evolving conditions. Although the functional H_2-sensing cascade has not yet been established in *Synechocystis* yet, we utilized the associated two-component system consisting of a histidine kinase and a response regulator to drive and modulate the expression of a *superfolder gfp* gene in *Escherichia coli*. This demonstrates that all components of the H_2-dependent signal cascade can be functionally implemented in heterologous hosts. Thus, this work provides the basis for the development of an intrinsic H_2 biosensor within a cyanobacterial cell that could be used to probe the effects of random mutagenesis and systematically identify promising genetic configurations to enable continuous and high-yield production of H_2 *via* oxygenic photosynthesis.

KEYWORDS

sensing and signaling, biotechnological hydrogen, regulatory hydrogenase, biosensor, synthetic biology, cyanobacteria

1. Introduction

The anthropogenic emission of greenhouse gases like carbon dioxide (CO_2) derived from the usage of fossil resources is regarded as the major driver of climate change. To tackle this issue, new approaches need to be supplied toward a CO_2-neutral society and economy. Molecular hydrogen (H_2) is generally believed to be an ideal candidate as a future energy carrier due to its high energy density and greenhouse gas emission-free usage. Industrially, H_2 is, however, still mainly obtained *via* steam reforming of natural gases, therefore relying on fossil resources and leading to a considerable greenhouse gas footprint (Howarth and Jacobson, 2021).

Biotechnological H_2 production using microorganisms as whole-cell biocatalysts offers the advantage of a sustainable process based on renewable resources. These biological H_2 formation routes encompass anaerobic fermentation using organic compounds as electron donors in, e.g., chemotrophic *Clostridium* and *Enterobacter* species or phototrophic sulfur and non-sulfur bacteria, as well as oxygenic photosynthesis using algae and cyanobacteria (Mahidhara et al., 2019). Approaches based on oxygenic photosynthesis appear most promising as they rely on electrons that have been obtained from light-dependent oxidation of water. Cyanobacteria are the only prokaryotes capable of this process. Great effort has been made to optimize H_2 production within cyanobacterial models such as the unicellular strain *Synechocystis* sp. PCC 6803 (hereafter referred to as *Synechocystis*). However, the breakthrough to enable continuous H_2 production in whole-cell cyanobacterial catalysts has not been achieved yet. Currently, it suffers from low yields and rates as well as the prototypical molecular oxygen (O_2) sensitivity of the enzymes involved in the formation of H_2, namely hydrogenases or nitrogenases. Hydrogenases are metalloenzymes that perform the reversible splitting of H_2 into protons and electrons. They are grouped based on the composition of their active site into nickel-iron [NiFe]-, iron–iron [FeFe]-, and iron [Fe]- or Hmd-hydrogenases (Lubitz et al., 2014). Previous studies tackled, for instance, the catalytic performance of H_2 production by introducing highly active, heterologous [FeFe]-hydrogenases into *Synechocystis* (Berto et al., 2011; Wegelius et al., 2018) or by fusing the endogenous [NiFe]-hydrogenase to photosystem I for a direct electron transfer from photosynthesis (Appel et al., 2020). The O_2 sensitivity has been addressed, for example, by introducing a heterologous O_2-tolerant [NiFe]-hydrogenase (Lupacchini et al., 2021). Another strategy enabling a continuous hydrogenase activity would be the spatial separation from O_2, e.g., through the encapsulation in synthetic microcompartments as demonstrated in *Escherichia coli* (hereafter referred to as *E. coli*) (Li et al., 2020). Moreover, metabolic engineering might target the redirecting of electron flows from competing pathways, like respiration and nitrate assimilation, to H_2 evolution (Baebprasert et al., 2011). Nevertheless, further research and alternative approaches are required to overcome the known as well as yet unknown limitations and to make photosynthesis-driven H_2 production amenable for biotechnological applications in the future (Bühler et al., 2021). In this regard, biosensors that respond to H_2 in an easily detectable way could help to enable, e.g., a systematic screening of mutant libraries and the selection of those that are beneficial for H_2 production.

Synthetic biosensors based on engineered bacterial cells, that respond to certain input stimuli with a desired output signal, can be designed by harnessing natural signal transduction systems to drive the expression of a reporter gene (Zhang et al., 2015; Ni et al., 2021). Also cyanobacteria have already been used as hosts to implement such cascades, e.g., for the intracellular sensing of heavy metals (Lacey et al., 2019; Patyi et al., 2021), O_2 (Immethun et al., 2016), or toluene (Inaba et al., 2018). Natural H_2-responsive systems were described in *Bradyrhizobium japonicum* (*B. japonicum*) (Black et al., 1994; van Soom et al., 1997, 1999), *Rhodobacter capsulatus* (*R. capsulatus*) (Dischert et al., 1999; Elsen et al., 2003), and *Cupriavidus necator* (also known as *Ralstonia eutropha*) H16 (hereafter referred to as *C. necator*) (Lenz et al., 1997; Lenz and Friedrich, 1998). The purple non-sulfur bacterium *R. capsulatus* has already been engineered to follow H_2 production in co-cultivated green algae (Wecker et al., 2011). However, such a co-cultivation approach impedes the use as a tool for efficient mutant screening and is not feasible in the case of cyanobacteria as most bacterial strains do not grow in cyanobacterial growth media. In the long-term, a cyanobacterial biosensor strain that directly responds to intracellularly evolved H_2 appears promising to use it as platform for a systematic optimization of H_2 production within the same cell.

Cupriavidus necator has become the model organism for H_2 oxidation in presence of O_2. As a true "*Knallgas*" bacterium it can utilize H_2 as sole electron donor and O_2 as terminal electron acceptor. For this purpose, it uses O_2-tolerant [NiFe]-hydrogenases. *C. necator* contains even four [NiFe]-hydrogenases that are O_2-tolerant, among them the soluble NAD^+-reducing (SH) and membrane-bound hydrogenase (MBH) as well as the regulatory hydrogenase (RH) (Lenz et al., 2015). The RH is a cytoplasmic enzyme and has a comparably simple structure composed of the two hydrogenase subunits, HoxB and HoxC (Pierik et al., 1998; Kleihues et al., 2000; Bernhard et al., 2001). The catalytic Ni-Fe center is coordinated by four cysteines to the protein matrix of the HoxC subunit (Winter et al., 2004). The active site iron carries three diatomic ligands, two cyanides (CN^-) and one carbon monoxide (CO) (Pierik et al., 1998). Incorporation of the active site into the RH requires a set of seven auxiliary maturases, which are encoded by the *hypA1B1F1CDEX* genes (Buhrke et al., 2001; Bürstel et al., 2016). The RH serves as H_2 sensor in combination with a two-component regulatory system and transmits the H_2 input signal *via* the histidine kinase HoxJ to the response regulator HoxA, which functions as transcriptional activator for hydrogenase gene expression (Zimmer et al., 1995; Lenz et al., 1997; Schwartz et al., 1998; Figure 1A). Unlike canonical two-component systems, transcriptional activation is mediated by the unphosphorylated form of HoxA, and HoxJ phosphorylates/inactivates HoxA in the absence of H_2 (Lenz and Friedrich, 1998). Notably, the phosphorylating activity of HoxJ is knocked out in its parental stain *C. necator* H16. To restore the native activity of HoxJ, a specific amino acid exchange is required, resulting in a functional HoxJ(S422G), also referred to as HoxJ* (Lenz and Friedrich, 1998). HoxJ* and the RH form the ternary H_2-sensing complex (Buhrke et al., 2004). Production of active RH has already been established in *E. coli* (Lenz et al., 2007), very recently even under aerobic conditions (Fan et al., 2022).

In this study, we introduced the H_2-sensing module of the *C. necator* RH, i.e., HoxB and HoxC, into *Synechocystis*. For this purpose, synthetic operons for the structural and accessory genes were designed for expression in *Synechocystis*. Heterologously produced and catalytically active RH was extracted from photoautotrophically grown cells that continuously evolve O_2. Furthermore, as proof of

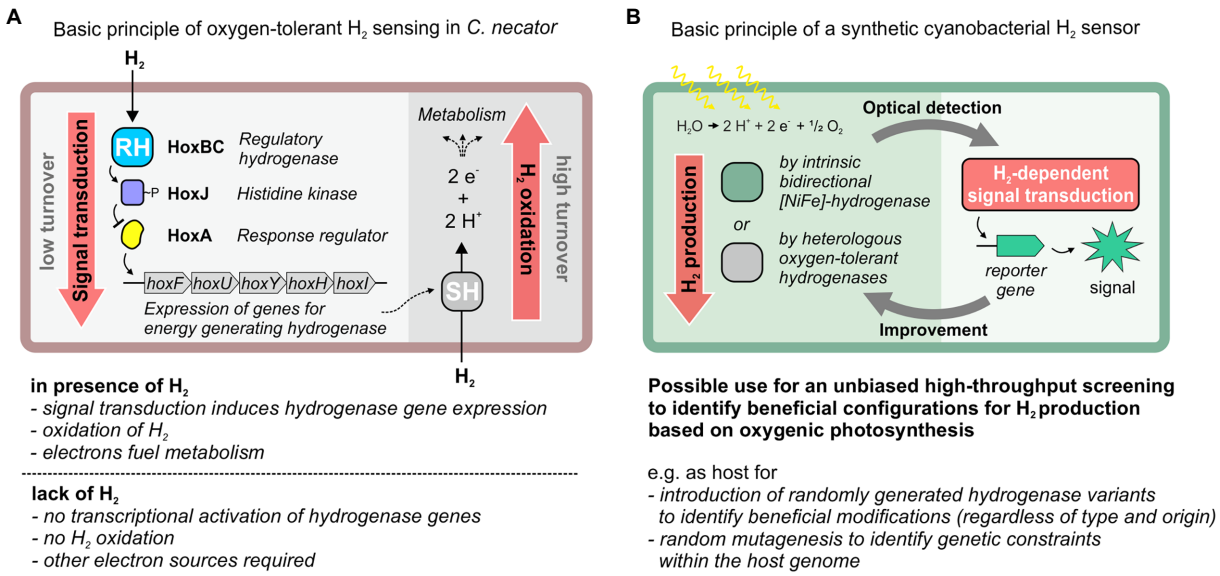

FIGURE 1
H_2-dependent signal transduction in *C. necator* and its potential utilization for the design of a cyanobacterial H_2 sensor strain that could be used as platform for the improvement of sustainable H_2 production *via* oxygenic photosynthesis. **(A)** H_2-responsive signal cascade in *C. necator* that enables the utilization of molecular hydrogen (H_2) when other (favored) energy sources are absent. Signal transduction involves the regulatory hydrogenase (RH) module HoxBC as H_2 sensor (with very low catalytic activity for H_2 oxidation), the histidine kinase HoxJ and the response regulator HoxA, which promotes transcription of genes for two oxygen-tolerant hydrogenases [a soluble NAD^+-reducing hydrogenase (SH) or a membrane-bound hydrogenase (MBH)]. These hydrogenases show high activity toward H_2 oxidation and are used for energy-conservation. **(B)** Potential application of this H_2-responsive signal cascade to drive reporter gene expression in cyanobacteria. This would enable optical detection of H_2 that was produced by other hydrogenases using electrons obtained by light-driven water oxidation. This screening platform could facilitate strain engineering with diverse strategies, e.g., improving hydrogenase properties by (repeated) rational design or random mutagenesis, toward an improved photosynthetic H_2 production. SH, soluble NAD^+-reducing hydrogenase; RH, regulatory hydrogenase.

concept, we introduced functional HoxJ* and HoxA into *E. coli* to modulate the expression of a reporter gene fused to a HoxA-responsive heterologous promoter. Our study provides the basis for further engineering of a cyanobacterial H_2 biosensor strain that might enable a systematic screening of genetic setups and the selection of those beneficial for H_2 production by the host-specific or other introduced hydrogenases (Figure 1B).

2. Materials and methods

2.1. Strains and culture conditions

Escherichia coli strains DH5α or JM109 were grown at 37°C either on agar-solidified LB medium or in LB liquid medium supplemented with 5 g L^{-1} NaCl under continuous shaking at 200 rpm. To select for the presence of certain plasmids the medium was supplemented with 100 μg mL^{-1} ampicillin, 35 μg mL^{-1} chloramphenicol, or 50 μg mL^{-1} spectinomycin. *C. necator* (obtained from the German Collection of Microorganisms and Cell Cultures, DSMZ) was grown at 37°C in LB liquid medium supplemented with 2.5 g L^{-1} NaCl under continuous shaking at 200 rpm. *Synechocystis* was cultivated in yBG11 (Shcolnick et al., 2007) liquid medium under continuous shaking at 150 rpm, or BG11 (Stanier et al., 1979) solidified with 1.5% (w/v) Bacto agar (Becton Dickinson) and supplemented with 3 g L^{-1} $Na_2S_2O_3$. The cyanobacterial growth media were buffered with 10–50 mM HEPES to pH 7.2. Photoautotrophic growth conditions were set to 30°C, ambient CO_2, constant light illumination with 50 μmol photons m^{-2} s^{-1}, and 75% (v/v) humidity. For the selection of mutants, the media were supplemented with 10 μg mL^{-1} chloramphenicol, 20 μg mL^{-1} spectinomycin, or 50 μg mL^{-1} kanamycin. A non-motile, glucose tolerant strain of *Synechocystis*, originally received from Martin Hagemann (Rostock University, Germany), was used as the wild type (WT). The mutant *Synechocystis(Δhox)* that is devoid of the endogenous [NiFe]-hydrogenase was obtained from Kirsten Gutekunst (Kassel University, Germany). In particular, *hoxEFUYH* (*sll1220-sll1226*) have been replaced by a kanamycin resistance cassette (Appel et al., 2020).

2.2. Construction of plasmids and recombinant strains

In silico work was performed using the software Geneious (Biomatters). Genetic constructs were generated through standard molecular cloning procedures and maintained on plasmids in *E. coli* DH5α. DNA processing and recombination were performed using FastDigest restriction endonucleases (Thermo Scientific), T4 DNA ligase (Thermo Scientific), and FastAP thermosensitive alkaline phosphatase (Thermo Scientific) following the manufacturer's instructions. The obtained constructs were verified by Sanger sequencing. Information about used oligonucleotides and plasmids is given in Supplementary Tables S1, S2.

The sequences for the design of synthetic operons encoding the H_2-sensing complex and the corresponding maturases were obtained from the megaplasmid pHG1 of *C. necator* (Schwartz

et al., 2003). The *hoxJ* sequence was modified to code for a variant exhibiting an amino acid substitution from serine to glycine at position 422, denoted as HoxJ* (Lenz and Friedrich, 1998). The *hoxC* sequence was altered to instead encode the variant HoxC(D15H) (Gebler et al., 2007). The gene sequences were codon-usage optimized for *Synechocystis* using the web-based tool JCat (Grote et al., 2005). The inducible promoters P_{rhaBAD} (Behle et al., 2020) and P_{nrsB} (Englund et al., 2016) were fused to the *hox* and *hyp* operons, respectively. The constructs were also equipped with unique restriction endonuclease sites flanking each operon, effective ribosome binding sites (RBS*, Heidorn et al., 2011) upstream of each ORF, as well as standardized transcription terminators (BioBrick BBa_B0015, Registry of Standard Biological Parts, 2003). Tag-encoding sequences were added to the 3′ ends of the open reading frames of *hoxB* (3xFLAG-tag), *hoxJ** (3xFLAG-tag) and *hypX* (Strep-tag), respectively. The synthetic operons were chemically synthesized (Eurofins Genomics) and provided on plasmids, denoted as pHox2 and pHyp. For the maintenance in *Synechocystis*, the synthetic operons were transferred, either separately or combined, into the plasmid pSHDY_P_{rhaBAD}::*mVenus*_P_{J23119}::*rhaS* (Behle et al., 2020) *via* *Xba*I/*Bcu*I and *Bcu*I/*Pst*I sites, respectively. Thereby, the P_{rhaBAD}::*mVenus* cassette was replaced. The resulting plasmids were named pHySe_Hox and pHySe_Hox_Hyp (Figure 2). The full sequences including annotations are provided in the Supplementary Data.

Plasmids of the pFO series and derivatives of pSB1A2_P_{trc1O} (Huang et al., 2010), all of which harbor various combinations of expression cassettes for the genes *hoxA*, *hoxJ** and *sfgfp*, were generated *via* the Gibson assembly procedure (Gibson et al., 2009). For this, respective sequences were amplified *via* PCR using primers with 5′ extensions to create homologous overhangs for the desired assembly with other DNA fragments. The P_{SH} promoter, including the 5′ untranslated upstream region of *hoxF* (Zimmer et al., 1995; Schwartz et al., 1998), was amplified from genomic DNA of *C. necator* using primer pair P17/P18. Together with the *sfgfp* reporter gene and a downstream BioBrick BBa_B0015 transcription terminator, which were generated through PCR with primers P16/Sam_102 from pSEVA351-*sfgfp* (Opel et al., 2022), it was used for the assembly of pFO6 utilizing *Kpn*I-treated pSEVA351 (Martínez-García et al., 2020) as vector. The *hoxA* gene was amplified from gDNA of *C. necator* using P25 that additionally contained the ribosome binding site BioBrick BBa_B0034 (Registry of Standard Biological Parts, 2003) as 5′ extension and P26 fusing a sequence encoding a hexahistidine-tag at the 3′ end of the gene. The linear fragment was used for the assembly with *Bcu*I-cut pSB1A2, yielding pSB1A2_P_{trc1O}-*hoxA*. pSB1A2_P_{trc1O}-*hoxA*^{D55A} has a substitution at

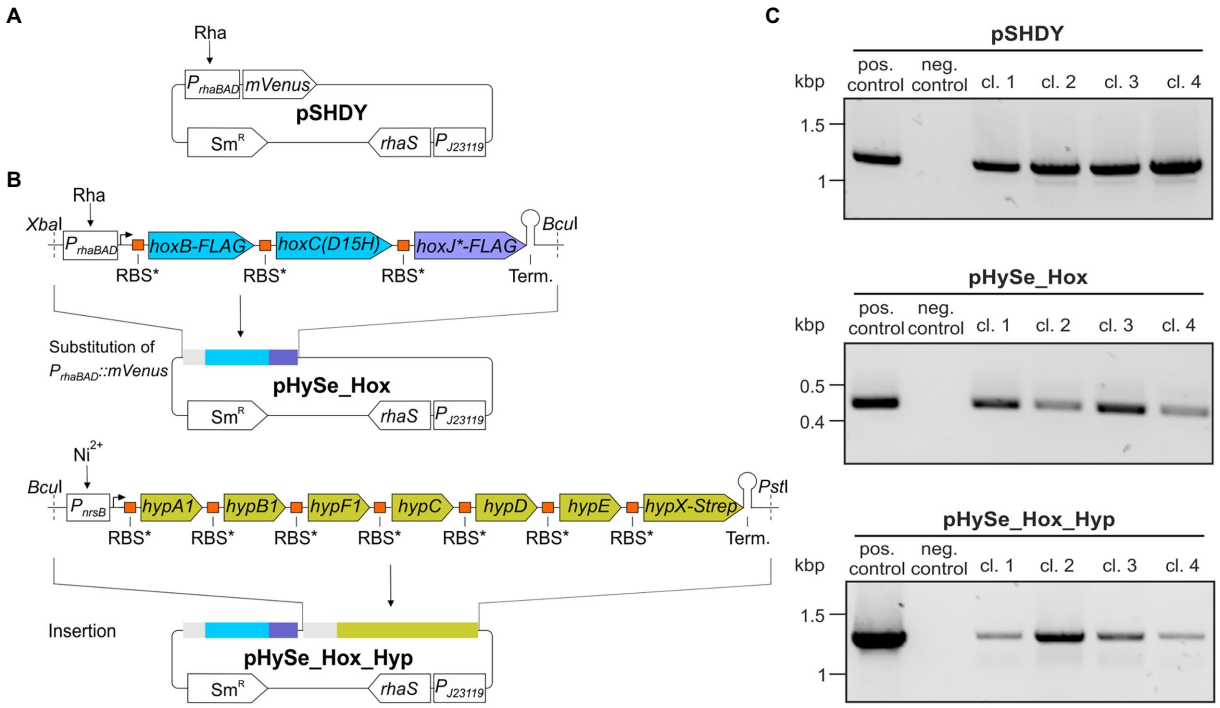

FIGURE 2
Scheme of synthetic operons for the expression of *C. necator* genes encoding the H_2-sensing HoxBCJ* complex and its associated maturases in *Synechocystis*. **(A)** Structure of the plasmid pSHDY_P_{rhaBAD}::*mVenus*_P_{J23119}::*rhaS* (herein referred to as pSHDY). It can be maintained in *Synechocystis*, contains a spectinomycin selection marker (Sm^R) for selection and enables L-rhamnose (Rha)-inducible reporter gene expression *via* P_{rhaBAD} by featuring a cassette for the constitutive expression of *rhaS* (Behle et al., 2020). **(B)** Structure of the pHySe plasmids. The *hox* and *hyp* genes are organized in synthetic operons with a transcription terminator (Term.) at the 3′ ends. Their polycistronic transcription is driven by P_{rhaBAD} and the nickel-ion (Ni^{2+})-responsive P_{nrsB} (Englund et al., 2016), respectively. Each gene is interconnected by a ribosome binding site, RBS* (Heidorn et al., 2011). **(C)** Verification of recombinant *Synechocystis* strains. A *Synechocystis* parental strain lacking the endogenous hydrogenase was transformed using either pSHDY, pHySe_Hox, or pHySe_Hox_Hyp. Four selected clones (cl.1–4) were analyzed by colony PCR with primers that specifically targeted regions of the particular vector, i.e., S17 and S31 for pSHDY, S22 and S31 for pHySe_Hox, as well as S30 and S31 for pHySe_Hox_Hyp. The respective pure plasmid DNA served as positive control (pos. control). A PCR reaction lacking a template was used as negative control (neg. control).

the codon coding for the amino acid at position 55 of HoxA from 5′-GAT-3′ (Asp) to 5'-GCC-3′ (Ala). It was generated with the *Dpn*I-digested product from an inverted PCR, taking pSB1A2_P_{trc1O}-*hoxA* as template and the primer pair P52/P53, as well as the homologous overhangs-supplying HoxA(D55A) double-stranded DNA fragment. The P_{trc1O}-*hoxA* and P_{trc1O}-$hoxA^{D55A}$ constructs were PCR-amplified using primers P49 and P50 from either pSB1A2_P_{trc1O}-*hoxA* or pSB1A2_P_{trc1O}-$hoxA^{D55A}$, thereby fused to a 3'sequence encoding a Strep-tag II, instead of the His-tag, to be each inserted into *Bcu*I-linearized pSEVA351, yielding pFO25 and pFO26, respectively. Analogously, these two synthetic gene constructs were inserted into *Bcu*I-cut pFO6, which resulted in pFO27 and pFO28, respectively. To obtain pFO45 and pFO46, pHox2 was first subjected to an inverted PCR using P83 and P84, thereby deleting $hoxB^{FLAG}$ and $hoxC^{D15H}$ as well as the particular upstream situated RBS*. This was followed by AQUA cloning, creating pHox5 that encodes the P_{rhaBAD}::$hoxJ^{*FLAG}$ cassette. The latter was excised by restriction with *Xba*I and inserted into *Xba*I-linearized pFO27 and pFO28, yielding pFO45 and pFO46, respectively. Sequences of the pFO series are provided in the Supplementary Data.

Synechocystis WT as well as *Synechocystis(Δhox)* parental cells were made electro-competent and transformed *via* electroporation as described previously (Brandenburg et al., 2021). Plasmid-harboring strains were selected on BG11 agar plates containing appropriate antibiotics. Plasmid presence was verified by colony PCR using suitable primers and the GoTaq MasterMix (Promega) according to the manufacturer's instructions. Recombinant *E. coli* JM109 strains were generated by electroporation of electro-competent cells *via* standard procedures.

2.3. RNA isolation and transcript analyses

For the isolation of RNA, *Synechocystis* cells were grown until reaching an OD_{750} ~ 0.8. The cultures were subsequently supplemented with final concentrations of 0.1% (w/v) L-rhamnose and 5 µM $NiSO_4$. After 24 h, cells were harvested by rapid vacuum filtration applying sterilized polyether sulfone filters (pore size 0.8 µm, PALL). RNA isolation was performed as described previously (Bolay et al., 2022). The RNA samples were treated with RNase-free DNase I (Thermo Scientific) according to the manufacturer's instructions. Afterwards, cDNA was generated by applying the high-capacity cDNA reverse transcription kit (Thermo Scientific) as given in the manufacturer's instructions. A total of ~0.4 ng cDNA were used as template for quantitative PCR. Amplification of specific regions within either the *rnpB* gene or *hypX* were performed using the GoTaq MasterMix (Promega) according to the manufacturer's instructions and primer pairs rnpB_114F/rnpB_226R and S30/P88, respectively (Supplementary Table S1).

2.4. Protein extraction and Western blots

Synechocystis cells were grown in presence of elevated CO_2 concentration of 2% (v/v) to an OD_{750} of ~2. To induce expression of the *hox* and *hyp* genes the medium was supplemented with final concentrations (f.c.) of 0.2% (w/v) L-rhamnose and 5 µM $NiSO_4$. In addition, 17 µM (f.c.) ferric ammonium citrate was added to foster hydrogenase maturation similar to previous reports (Lupacchini et al., 2021). Samples were collected by centrifugation after 24 and 48 h. Cells were resuspended in 750 µL TBS lysis buffer (100 mM Tris, 150 mM NaCl, 1 mM PMSF, pH 7.5) and transferred to 2 mL Precellys tubes (Bertin), together with a mixture of glass beads (Sartorius) of 0.09–0.15, 0.17–0.18, and 0.5 mm diameter. Cell disruption was performed using a Precellys Evolution homogenizer (Bertin) equipped with a Cryolys cooling system (Bertin) for 4 × 30 s at 10.000 rpm with 30 s interim breaks for cooling. The samples were subsequently separated in supernatant (soluble extract) and sediment (crude extract) by centrifugation and subjected to protein concentration determination using a Bradford dye reagent ready-to-use solution (Thermo Scientific) according to the manufacturer's instructions. Cell suspensions of recombinant *E. coli* JM109 strains were analogously treated to obtain soluble protein extracts. Those cells were beforehand cultivated as described for GFP fluorescence determination, but using 0.5% (w/v) D-glucose instead of glycerol. Protein separation was performed *via* SDS-PAGE using a total amount of 20 µg protein for each sample. For immunodetection *via* Western blots the separated proteins were transferred to nitrocellulose membranes of 0.45 µm pore size (GVS), followed by hybridization with either a Strep-Tactin horse radish peroxidase (HRP) (IBA Lifesciences GmbH) or a monoclonal ANTI-FLAG M2-Peroxidase conjugate (Sigma-Aldrich) according to the manufacturer's instructions. Chemiluminescence was detected by using the substrate solutions WesternSure PREMIUM Chemiluminescent (LI-COR) or WesterBright ECL (advansta) and the Fusion FX7 EDGE V0.7 imaging system (VILBER), following the manufacturer's instructions.

2.5. Hydrogenase activity assays

Pre-cultivation of *Synechocystis* was performed as described above and expression of heterologous genes was induced by 0.1% (w/v) L-rhamnose and/or 2.5 µM $NiSO_4$. After 48 h, cells were harvested and disrupted analogously but using an alternative lysis buffer (5% (v/v) glycerol, 50 mM KPO_4, 1 mM PMSF, pH 8). A *Synechocystis* strain harboring the *hoxFUYHW* genes encoding the SH from *C. necator* (Lupacchini et al., 2021) served as control. Approximately 400 µg of soluble proteins were separated *via* native PAGE and subjected to in-gel staining as previously described (Lupacchini et al., 2021) with few modifications. These concerned the supplementation of 90 µM phenazine methosulfate, additionally to 800 µM NAD^+ and 60 µM nitro blue tetrazolium (NBT) in an H_2-saturated activity buffer (50 mM Tris, pH 8). Furthermore, the incubation time was increased from ~0.5 h to ~2.5 h. Hydrogenase activity, i.e., the release of electrons from H_2 oxidation, is indicated *via* a step-wise reduction of the electron transfer mediators NAD^+ and/or phenazine methosulfate and the colorimetric dye nitro blue tetrazolium, which finally results in a visible precipitation of formazan (Ponti et al., 1978). The gels were subsequently decolorized from the remaining loading dye in activity buffer overnight. Afterwards, presence of HoxB and HoxJ* proteins was confirmed by blotting the same polyacrylamide gel and hybridizing the membrane with antibodies against the attached 3xFLAG-tag as described above.

2.6. GFP fluorescence determination

Recombinant *Synechocystis* strains containing plasmids pFO25 (negative control), pFO6 (P_{SH}::*sfgfp*), pFO27 (P_{SH}::*sfgfp* + P_{trc1O}::*hoxA*), or pFO28 (P_{SH}::*sfgfp* + P_{trc1O}::*hoxA*^{D55A}) were analyzed *in vivo* regarding GFP fluorescence. The detection was performed as described previously (Opel et al., 2022). GFP fluorescence determination in *E. coli* JM109 was conducted for recombinant strains harboring the following plasmids: pSEVA351 (negative control), pFO27 (P_{SH}::*sfgfp* + P_{trc1O}::*hoxA*), pFO45 (P_{SH}::*sfgfp* + P_{trc1O}::*hoxA* + P_{rhaBAD}::*hoxJ**), and pFO46 (P_{SH}::*sfgfp* + P_{trc1O}:: *hoxA*^{D55A} + P_{rhaBAD}::*hoxJ**). Single colonies from selective LB agar plates were picked to inoculate liquid LB medium pre-cultures that were grown for ~18 h at 37°C. 1% (v/v) of these suspensions were taken to inoculate second pre-cultures using M9* medium, supplemented with 0.001% (w/v) thiamine, 2 mM $MgSO_4$, 0.4% (v/v) glycerol, US* trace elements solution, and buffered to pH 7.2. The M9* pre-cultures were incubated for ~24 h at 37°C. A volume of 100 µL of these cell suspensions were added to 10 mL M9* medium in baffled shake flasks, and the cells were further cultivated at 30°C instead, due to the temperature sensitivity of HoxA (Zimmer et al., 1995). These main cultures were supplemented with 10 µM IPTG and/or 0.2% L-rhamnose after 8 h, followed by another 16 h of cultivation. For GFP fluorescence determination, samples were diluted to an OD_{600} of ~0.5 with TBS buffer (100 mM Tris, 150 mM NaCl, pH 7.5) in a final volume of 1,200 µL. Technical triplicates (each 200 µL) were transferred into an opaque black flat microtiter 96-well-plate (Nunc), followed by fluorescence measurements using an Infinite 200 PRO microplate reader (Tecan; gain: 123, integration time: 20 µs, excitation bandwidth: 9 nm, emission bandwidth: 20 nm, z-position: 2000 µm, 25 flashes) and excitation/emission wavelengths of 485 nm/520 nm, respectively. Furthermore, the same sample was used to measure the absorption at 600 nm in a transparent flat microtiter 96-well-plate (Nunc), also using the Infinite 200 PRO microplate reader (bandwidth: 9 nm, 25 flashes). The blank of the TBS buffer background was subtracted from the values detected for the cell suspensions. The fluorescence intensities were finally normalized by division of respective OD_{600} values.

3. Results

3.1. Design of customized operons for the expression of a H_2-sensing complex in a cyanobacterial host

To functionally produce the RH of *C. necator* in *Synechocystis*, we rationally designed two operons *in silico* that were generated *via* chemical synthesis and inserted it into a vector that can be maintained in *Synechocystis* (Figures 2A,B). In case of the active site-containing HoxC subunit, we made use of the amino acid exchange variant HoxC(D15H), which, in contrast to the native protein, supports H_2-dependent growth of *C. necator* at an O_2 level of up to 10% (Gebler et al., 2007). The resulting gene cluster for the biosynthesis of the H_2-sensing complex comprised *hoxB, hoxC*^{D15H}, and *hoxJ**. For proper maturation of the catalytic [NiFe] center we also utilized the corresponding accessory genes *hypA1B1F1CDEFX* (Buhrke et al., 2001; Bürstel et al., 2016). In *C. necator*, these genes are organized in an operon structure with partially overlapping open reading frames (Schwartz et al., 2003). To ensure their correct expression in the cyanobacterial target organism, we altered the spatial organization by linking each *hox* and *hyp* gene by an artificial spacer region, resulting in separate translational units (Figure 2B). Furthermore, the synthetic ribosome binding site RBS* which functions in *Synechocystis* (Heidorn et al., 2011), was introduced upstream of every single gene to enable efficient translation initiation. For facile detection of the proteins, sequences encoding either a 3xFLAG-tag or a Strep-tag were fused to *hoxB*, *hoxJ**, and *hypX* (Figure 2B). The *hoxJ** and *hypX* genes were chosen because they are the dorsal genes of each particular operon. Detection of both proteins is considered representative of upstream gene expression.

All protein-coding sequences were codon-usage optimized for translation in *Synechocystis*. To drive the polycistronic transcription of *hox* and *hyp* gene clusters, we used the L-rhamnose-inducible promoter P_{rhaBAD} from *E. coli* (Behle et al., 2020) and the nickel ion (Ni^{2+})-dependent promoter P_{nrsB} of *Synechocystis* (Englund et al., 2016), respectively. Thus, this setup permits a selective induction as well as a tight and tunable transcription of the synthetic gene constructs P_{rhaBAD}::*hoxB*^{FLAG}*C*^{D15H}*J**^{FLAG} (*hox* operon) and P_{nrsB}::*hypA1B1F1CDEFX*^{Strep} (*hyp* operon) in *Synechocystis*. Moreover, both operons were equipped with insulating transcription terminators at their 3′ end.

The constructs were inserted into the replicative vector pSHDY_P_{rhaBAD}::*mVenus* _P_{J23119}-*rhaS* (hereafter referred to as pSHDY), which encodes the heterologous transcriptional regulator RhaS that enables rhamnose-inducible gene expression in *Synechocystis* (Behle et al., 2020). The P_{rhaBAD}::*mVenus* cassette present in this pSHDY construct (Figure 2A) was replaced with the synthetic *hox* operon resulting in the plasmid pHySe_Hox. Subsequently, the *hyp* operon was inserted downstream to obtain the plasmid pHySe_Hox_Hyp (Figure 2B). The resulting plasmids pHySe_Hox and pHySe_Hox_Hyp, as well as the precursor construct pSHDY, were used individually for the transformation of *Synechocystis*. For transformation, a strain devoid of the endogenous [NiFe]-hydrogenase, designated *Synechocystis(Δhox)* (Appel et al., 2020), was used to prevent subsequent cross-reactions with the RH activity. Plasmid presence was verified in all obtained clones (Figure 2C).

3.2. The genes encoding the H_2-sensing complex are expressed in *Synechocystis*

C-terminal linkage with 3x-FLAG (HoxB & HoxJ*) or Strep-tags (HypX), enabled protein detection by commercially available antibodies targeting the corresponding tag. The *Synechocystis* strain harboring pHySe_Hox_Hyp was grown in the presence of L-rhamnose and Ni^{2+} to trigger *hox* and *hyp* gene expression, respectively (see section "Materials and methods" for specific inducer concentrations). To confirm heterologous gene expression, we performed immunoblotting to detect HoxB and HoxJ* using protein extracts of samples collected 24 and 48 h after induction. In addition to the crude extract, we also analyzed the soluble protein fraction obtained by centrifugation. In fact, distinct bands were detected, and their intensity increased according to the induction time. The bands represent HoxB-FLAG and HoxJ*-FLAG, as no signal was detected in the same size

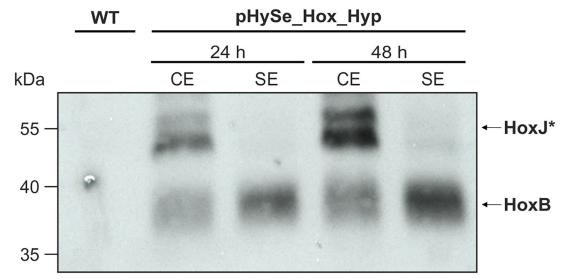

FIGURE 3
HoxJ* and HoxB from *C. necator* are synthesized in *Synechocystis*. Western blot for the detection of HoxJ*-FLAG and (~54kDa) HoxB-FLAG (~40kDa) fusion proteins. Samples were taken from *Synechocystis(Δhox)* that harbored the plasmid pHySe_Hox_Hyp, 24 and 48h after induction with 0.2% (w/v) L-rhamnose and 5μM $NiSO_4$. Protein extract of a *Synechocystis* wild type (WT) strain served as negative control. CE, crude extract; SE, soluble extract.

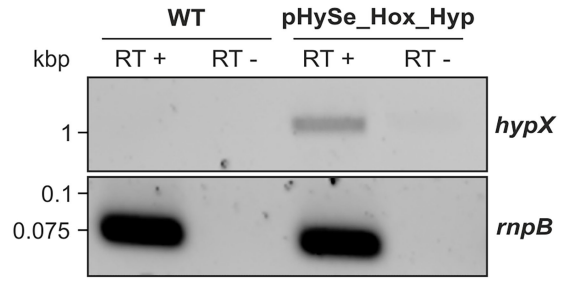

FIGURE 4
The polycistronic *hyp* mRNA is being transcribed in *Synechocystis*. PCR amplification of *hypX* or the housekeeping gene *rnpB* from cDNA that has been prepared from RNA isolates of *Synechocystis* wild type (control) or a strain harboring pHySe_Hox_Hyp. Samples in lanes denoted 'RT +' were treated with reverse transcriptase beforehand, while those indicated 'RT –' were not.

range in the *Synechocystis* wild-type strain (WT) (Figure 3). Consistent with the expected cytoplasmic localization of the RH, a strong signal for HoxB-FLAG was observed in the soluble extracts. However, the signal associated with HoxJ*-FLAG, which is also thought to be soluble, was predominantly present in the crude extract indicating partial protein misfolding or membrane association. Nevertheless, even though the signal was quite weak, a significant part was also found in the soluble fraction, in particular 48 h after induction. Thus, both fusion proteins were specifically detected, confirming the expression of the corresponding genes.

In case of HypX-Strep, however, no specific signal was observed in cells containing pHySe_Hox_Hyp (not shown), indicating either no translation or protein instability. To verify that the synthetic *hyp* operon is at least transcribed, we extracted total RNA from the same strain and performed classical reverse transcriptase (RT)-PCR targeting *hypX*. The reversely transcribed copy DNA (cDNA) for *hypX* was only detected in the *Synechocystis* strain containing pHySe_Hox_Hyp but not in the WT (Figure 4). That the band obtained is indeed a result of cDNA amplification of a *hypX* transcript was verified by a parallel RNA sample that was not treated with reverse transcriptase and consequently showed no bands for *hypX* and the housekeeping gene *rnpB*. As *hypX* is situated at the 3′ end of the synthetic gene cluster, we assume, that transcription of the upstream situated *hyp* genes also occurred. However, this analysis did not confirm the translation of the *hyp* gene transcripts. Nevertheless, sufficient synthesis of the maturation apparatus for the RH could be assumed, as indicated by the subsequent analysis (see below).

3.3. The H_2-sensing complex is active when matured in *Synechocystis* cells growing photoautotrophically

To investigate if the recombinant gene expression indeed results in the formation of active H_2-sensing RH, we analyzed the H_2 oxidation activity in soluble extracts of photoautotrophically grown cells of *Synechocystis* strains harboring the plasmids pSHDY, pHySe_Hox, and pHySe_Hox_Hyp. To this end, we used an in-gel activity assay under an H_2 atmosphere (Buhrke et al., 2004), which has also been used recently to confirm the activity of hydrogenases in *Synechocystis* (Lupacchini et al., 2021). The recombinant strains were cultivated in the presence of different inducer combinations to achieve independent expression of the *hoxBCJ** genes (L-rhamnose) and the *hypA1B1F1CDEX* operon (Ni^{2+}). Soluble protein extracts were prepared and subjected to native polyacrylamide gel electrophoresis. Strikingly, in-gel H_2 oxidation activity was detected only for *Synechocystis*(pHySe_Hox_Hyp) induced with both L-rhamnose and Ni^{2+} (Figure 5A, In-gel staining panel). The resulting activity bands were located at the same positions as the bands in the immunoblot analysis (also based on the native gel), showing HoxB-FLAG in complex with HoxC as the hydrogenase core module, and potentially HoxJ*-FLAG (Figure 5A, Western Blot panel). As the production of both HoxB-FLAG and HoxJ*-FLAG has been confirmed by a previous Western blot (Figure 3), a separate detection of both proteins has not been performed in this case. No H_2 oxidation activity was detected in the corresponding native gel, when the Hyp proteins required to produce catalytically active RH were absent, either due to the lack of the *hyp* genes (in case of strain pHySe_Hox) or the inducer Ni^{2+} (in case of strain pHySe_Hox_Hyp) (Figure 5A). Thus, expression of both the *hox* and *hyp* gene clusters is required to obtain detectable H_2 oxidation activity for the H_2-sensing RH in *Synechocystis*. While most [NiFe]-hydrogenases are inactivated by traces of O_2 (Shafaat et al., 2013), the RH activity has shown to be O_2-tolerant (Ash et al., 2015). Remarkably, the RH activity was observed in extracts of O_2-evolving photoautotrophically grown *Synechocystis* cells, indicating that both the maturation process and the catalytic activity of the RH occurred in the presence of O_2.

The RH activity in *Synechocystis* is comparatively low, as demonstrated by the long incubation time of ~2.5 h required to obtain detectable bands derived from H_2-dependent NBT reduction in the activity gel. A *Synechocystis* control strain containing the highly active SH from *C. necator* (Lupacchini et al., 2021) showed significantly stronger signal intensities after ~0.5 h already. The low signal strength of the RH in *Synechocystis* could not be increased by enhanced L-rhamnose levels, suggesting a saturation at 0.05% (w/v) rhamnose and consequently no limitation of the RH structural proteins (Figure 5B). Altogether, these results demonstrate for the first time the functional production of a recombinant regulatory hydrogenase with low H_2 turnover activity in a cyanobacterium.

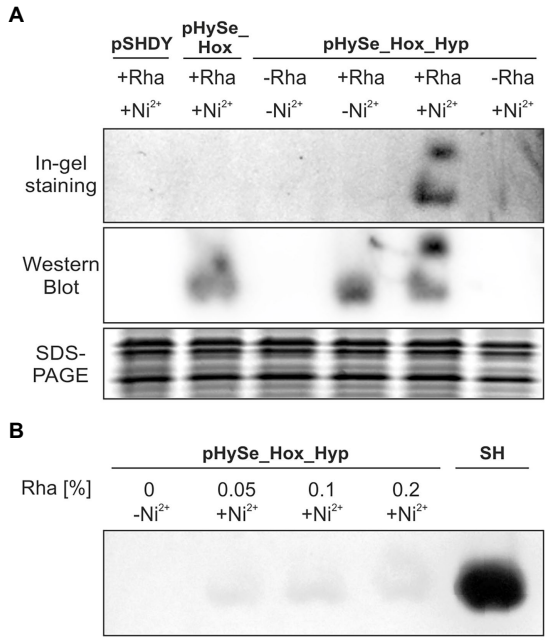

FIGURE 5
RH-mediated H_2 oxidation in *Synechocystis*. **(A)** Recombinant *Synechocystis*(Δ*hox*) strains lacking the endogenous hydrogenase and containing either pSHDY, pHySe_Hox, or pHySe_Hox_Hyp were grown photoautotrophically under aerobic conditions in the presence of the inducers, i.e., 0.1% (w/v) L-rhamnose (Rha) and/or 2.5μM Ni^{2+} (indicated by +/−). Soluble proteins were separated *via* native PAGE and subjected to in-gel H_2-dependent formazan staining to localize the H_2-sensing complex (upper panel), as well as immunological detection of the HoxB-FLAG and HoxJ*-FLAG proteins (middle panel). A denaturing SDS-PAGE was conducted as loading control (lower panel). **(B)** In-gel activity of cells incubated with the given inducers using different amounts of Rha in % (w/v). A *Synechocystis* strain recombinantly producing the highly active NAD^+-reducing soluble hydrogenase from *C. necator* (SH) served as positive control.

3.4. Transcriptional regulation and its modulation by the associated two-component system in a heterologous host

The final goal is to couple the functional RH with the associated kinase HoxJ* and the cognate response regulator HoxA to establish an H_2-sensing signal transduction cascade in *Synechocystis*. In *C. necator*, HoxA positively controls the transcription of the genes encoding the SH and the MBH through binding to the promoters P_{SH} and P_{MBH}, respectively (Zimmer et al., 1995; Schwartz et al., 1998). The HoxA activity is modulated by HoxJ*-mediated phosphorylation, with transcriptional activation by HoxA in its non-phosphorylated state (Lenz and Friedrich, 1998). To drive HoxA-mediated gene expression in *Synechocystis*, the *hoxA* gene was set under control of the IPTG-inducible P_{trc1O} promoter, which has been shown to provide sufficient constitutive expression due to the lack of the *lac* repressor LacI in the cyanobacterial host (Huang et al., 2010). Furthermore, a *sfgfp* reporter gene encoding the superfolder green fluorescent protein (Pédelacq et al., 2006) (hereafter referred to as GFP) was fused to the HoxA-dependent promoter P_{SH} (Figure 6A). The synthetic gene constructs P_{SH}::*sfgfp* and P_{trc1O}::*hoxA* were introduced into *Synechocystis* WT, either separately or combined on a replicative plasmid. However, no significant GFP fluorescence beyond background activity was detected in any strain carrying the reporter gene construct alone or in combination with P_{trc1O}::*hoxA* (Figure 6B). An inactivation of HoxA by unspecific phosphorylation could be excluded because the GFP fluorescence was similar in a reporter strain containing the phosphorylation-insensitive variant HoxA(D55A) instead of HoxA. This variant cannot be phosphorylated at the crucial aspartate at position 55 and has been shown to be always active (Lenz and Friedrich, 1998). Although several attempts have been made to improve *hoxA* expression, e.g., by using different promoters as well as codon-usage optimized gene variants, HoxA-dependent *sfgfp* expression has not yet been achieved in *Synechocystis* (not shown).

However, to validate the general functionality of our genetic constructs and to establish HoxA and HoxJ*-dependent gene expression in a heterologous system, we introduced the same replicative plasmids harboring P_{trc1O}::*hoxA* and P_{SH}::*sfgfp* into *E. coli* JM109. As this host contains LacI, *hoxA* expression is IPTG-inducible. Indeed, a ~5-fold higher GFP fluorescence was detected in cells grow in presence of IPTG compared to cells grown in absence of IPTG (Figure 6C). Under non-induced conditions, the GFP fluorescence was similar to the background autofluorescence of control cells harboring the empty vectors. Thus, in contrast to *Synechocystis*, the expected activity of HoxA to promote reporter gene expression *via* the P_{SH} promoter was confirmed in *E. coli*.

Moreover, another synthetic gene construct encoding HoxJ* was included in the study to demonstrate the modulation of HoxA activity. HoxJ*-mediated phosphorylation of HoxA is expected to inactivate the response regulator (Lenz and Friedrich, 1998; Buhrke et al., 2004). A gene encoding HoxJ* carrying a C-terminal FLAG-tag ($HoxJ^{*FLAG}$) was set under control of the L-rhamnose-inducible promoter P_{rhaBAD}, and the resulting plasmid was introduced into the strain already harboring the Ptrc1O::*hoxA* construct. In general, this promoter could also be used in *Synechocystis* (Behle et al., 2020). Again, IPTG-induced *hoxA* gene expression resulted in GFP fluorescence. Remarkably, a significant decrease in GFP fluorescence was observed in cells that were grown in presence of both IPTG and L-rhamnose (Figure 6C). This correlates well with the exclusive detection of HoxJ* in protein extracts of cells grown in presence of L-rhamnose and carrying the respective gene construct (Figure 6D). The GFP level decreased to 50% compared to values obtained with strains either lacking the *hoxJ** gene or that do not sufficiently express *hoxJ** due to the absence of L-rhamnose. HoxA(D55A), which cannot be inactivated by HoxJ* through phosphorylation (Lenz and Friedrich, 1998), was again included as a control. As expected, a higher reporter signal was obtained with the corresponding strain in the presence of both IPTG an L-rhamnose than with the strain expressing wild-type *hoxA* as well as *hoxJ** (Figure 6C). This result also suggests that the response regulator in *E. coli* JM109 is not inactivated by unspecific phosphorylation. Thus, for HoxA, an effective ~30% reduction in reporter signal strength was achieved by HoxJ* compared to HoxA(D55A), demonstrating the desired modulation. Moreover, the HoxJ*-mediated decrease of HoxA-dependent GFP fluorescence was further tunable by different amounts of L-rhamnose (Figure 6E).

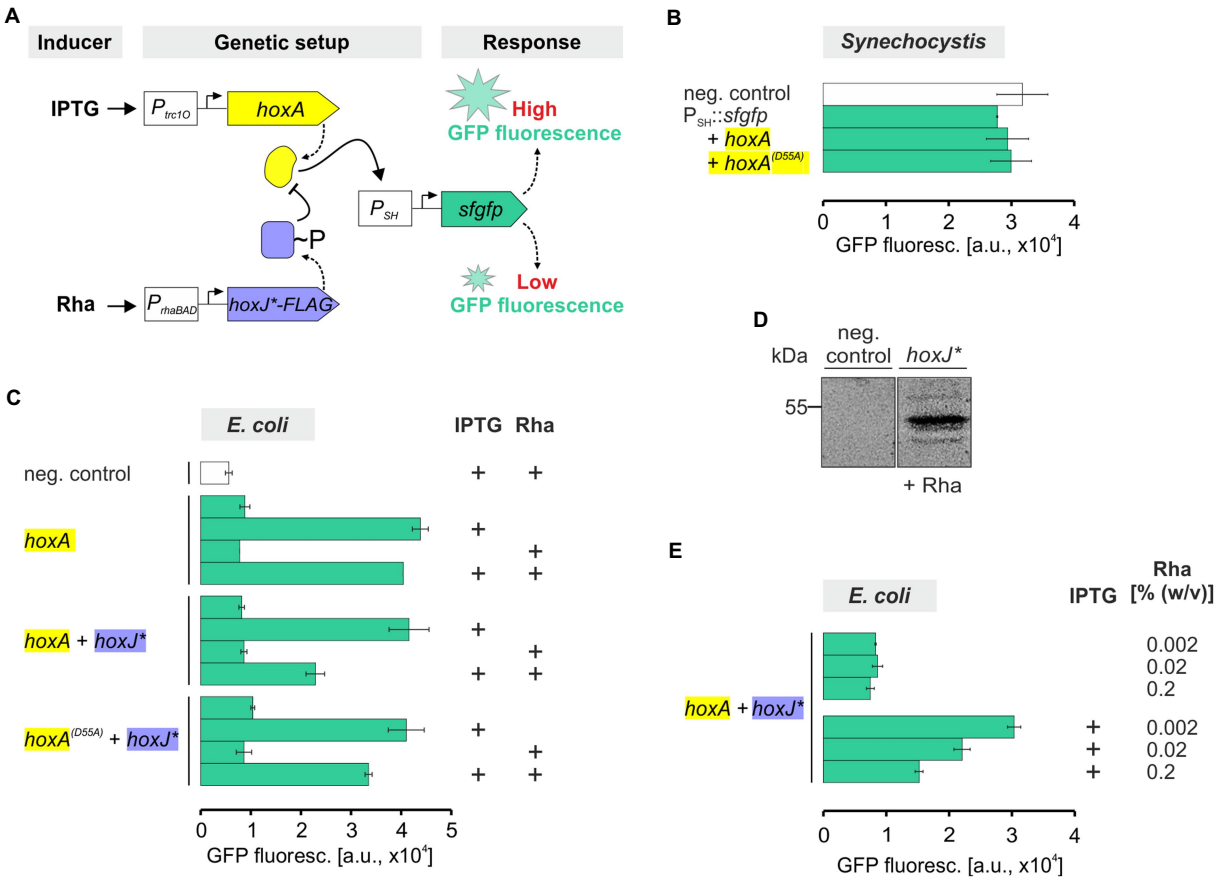

FIGURE 6
GFP fluorescence of *E. coli* and *Synechocystis* representing transcriptional regulation *via* HoxA and its modulation by HoxJ*. **(A)** Genetic setup of the generated strains. Expression of *hoxA* was driven by the IPTG-dependent *lac* promoter derivative P_{trc1O} (Huang et al., 2010). The expression of the *hoxJ** gene, fused to a FLAG-tag encoding sequence, was performed using the L-rhamnose (Rha)-inducible promoter P_{rhaBAD} (Behle et al., 2020). HoxA-dependent expression of a reporter was achieved by fusion of a *sfgfp* gene to the cognate promoter of the NAD^+-reducing, soluble hydrogenase gene cluster of *C. necator*, P_{SH} (Zimmer et al., 1995; Schwartz et al., 1998). HoxJ* in turn modulates HoxA-dependent transcriptional regulation (Lenz and Friedrich, 1998). **(B,C,E)** Whole-cell GFP fluorescence of recombinant *Synechocystis* or *E. coli* JM109 strains harboring the synthetic P_{SH}::*sfgfp* reporter gene construct in combination with different genetic information shown in panel **A**. Treatment with inducers IPTG and/or Rha prior to the measurements is indicated by '+.' Detected fluorescence was normalized to OD_{750} (panel **B**) or OD_{600} (panels **C** and **E**) and is given in arbitrary units (a.u.). The particular negative controls (neg. control) contained a plasmid without reporter gene cassette. Data are the mean±SD of at least two biological replicates (clones) each measured in technical triplicates. **(D)** Western blot confirming the presence of HoxJ*-FLAG (~54kDa) in soluble protein extracts of *E. coli* JM109 harboring a P_{rhaBAD}::$hoxJ^{*FLAG}$ cassette versus an empty vector neg. control.

4. Discussion

H_2-based signal transduction cascades are considered being widespread, as the corresponding genes have been detected in many available genomes and metagenomes (Greening et al., 2016). Nevertheless, the biochemical and molecular mechanisms of H_2 sensing have been studied in only a few representative bacteria to date. Ever since, these systems have been engineered and coupled with reporters. For example, *C. necator* has been engineered regarding a HoxBCJA-dependent expression of *lacZ* encoding β-galactosidase as reporter (Lenz and Friedrich, 1998). In principle, such recombinant strains could be utilized for the detection of H_2 synthesized by other microbes, e.g., by co-cultivation or agar overlay assays on petridish basis (Wecker and Ghirardi, 2014). Similar to *C. necator*, the H_2-sensing system of *R. capsulatus* consists of four proteins: HupUV which form the H_2-sensing hydrogenase, the histidine kinase HupT, and the transcriptional regulator HupR that finally activates expression of an energy-generating uptake hydrogenase (Dischert et al., 1999; Elsen et al., 2003). This system has also been developed into a biosensor to screen large libraries of H_2-producing nitrogenase variants in *R. capsulatus* directly (Barahona et al., 2016). Moreover, the corresponding strain has also been used as whole-cell biosensor to track H_2 production in co-cultivated green algae (Wecker et al., 2011). However, in the case of cyanobacteria, such a co-cultivation approach is not feasible since most bacterial strains do not grow in cyanobacterial growth media lacking an organic carbon source.

In order to monitor H_2 evolution within cyanobacterial cells, however, a signal transduction cascade must be transferred to the corresponding strain. This would allow *in vivo* screening of H_2 evolution and the optimization of strains carrying, e.g., alternative hydrogenases (Figure 1). For example, Wecker et al. used engineered *R. capsulatus* strains that monitor H_2 by reporter fluorescence to track

the activity of a recombinant H_2-evolving hydrogenase from *Clostridium acetobutylicum*. Only low H_2 production has been detected, but the system potentially enables further screening and hydrogenase evolution approaches (Wecker et al., 2017). We successfully implemented two parts of the four-part H_2-responsive signal transduction cascade from *C. necator* in the cyanobacterium *Synechocystis*. This is considered as a first step toward a synthetic cyanobacterial H_2 biosensor that could be used analogously to previous reports (Wecker et al., 2017). A potential application would be, for instance, to optimize the H_2 evolution activities of heterologously produced O_2-tolerant, energy-converting hydrogenases, one of which was successfully implemented in *Synechocystis* recently (Lupacchini et al., 2021).

According to the current model, H_2-sensing requires continuous H_2 activation, i.e., H_2 binding, H_2 cleavage, as well as the corresponding proton and electron transfer (Lenz et al., 2015). Protein–protein complex formation with HoxJ* is if course also required (Buhrke et al., 2004; Löscher et al., 2010). Thus, continuous H_2 oxidation is a prerequisite for H_2-sensing. While the H_2 oxidation activity of RH itself has been shown to be O_2 insensitive (Buhrke et al., 2005), the signal transduction process is sensitive to high O_2 levels. In this study, we therefore used a variant of the RH large subunit with an amino acid exchange near the active site, HoxC(D15H) (Gebler et al., 2007). The turnover rate of the native RH is almost two orders of magnitude lower than that of energy-conserving standard [NiFe]-hydrogenases (Bernhard et al., 2001). The H_2 oxidation activity of the HoxBC(D15H) variant is indeed another two orders of magnitude lower, which explains the weak signals observed in the in-gel activity assay (Figure 5). However, it mediated native-like H_2 signal transduction *in vivo* and supported H_2-dependent growth of *C. necator* even at 10% O_2, where signal transduction by the native RH was shown to be impaired (Gebler et al., 2007). The active-site containing HoxC subunit of RH also lacks a C-terminal extension that is typical for standard [NiFe]-hydrogenases and is proteolytically cleaved after insertion of the catalytic center (Kleihues et al., 2000). The comparatively simple structure, the lack of need for proteolytic quality control, the low H_2 consumption, and of course the O_2-tolerant H_2-sensing ability make the RH an ideal candidate for a synthetic cyanobacterial H_2 biosensor.

Here, we demonstrated that the recombinant RH in *Synechocystis* is synthesized in a catalytically active form, as evidenced by the H_2 oxidation activity detected in protein extracts. Moreover, the cell extracts were obtained from photoautotrophically grown, O_2-evolving cells. This shows that both the biocatalyst RH and its maturation machinery function properly under aerobic conditions in *Synechocystis*. In addition to the structural *hox* genes, the co-expression of the accessory *hypA1B1F1CDEX* genes of *C. necator* was required to achieve active RH. *Synechocystis* contains six endogenous Hyp proteins, likewise denoted HypA-F, that are responsible for the maturation of the bidirectional [NiFe]-hydrogenase of this organism (Hoffmann et al., 2006). However, *Synechocystis* lacks a homolog of the *C. necator* HypX. Based on our results, we cannot conclude whether all Hyp components or rather a reduced set from *C. necator* are necessary to achieve H_2 oxidation activity of the RH in *Synechocystis*. The *Synechocystis* Hyp proteins may at least partially compensate for the maturation of O_2-tolerant [NiFe]-hydrogenases from *C. necator* in the absence of the corresponding heterologous assembly apparatus. Notably, the NAD^+-reducing [NiFe]-hydrogenase from *C. necator* was functionally produced in *Synechocystis*, without co-expression of the associated *hyp* genes (Lupacchini et al., 2021). However, it has been suggested that the *C. necator* Hyp proteins may be required for full SH activity as they have an amino acid sequence identity with the *Synechocystis* homologs of only 50–67%. Furthermore, HypX is required for aerobic maturation of the [NiFe]-hydrogenases of *C. necator* (Bürstel et al., 2016; Schulz et al., 2020).

The transfer of signal-responsive components into heterologous hosts usually includes a promoter and the associated transcriptional regulator (Fernandez-López et al., 2015; Sonntag et al., 2020; Ni et al., 2021). Our objective was to transfer the H_2-sensing module (HoxBC) and the associated two-component regulatory system (HoxJ* and HoxA) to a cyanobacterial species to establish an H_2-dependent transcriptional response. This has not yet been achieved in *Synechocystis*, presumably due to the absence of the minor sigma factor σ^{54} in cyanobacteria (Riaz-Bradley, 2019), which is required for the transcriptional activation of P_{SH} and P_{MBH} in *C. necator* (Zimmer et al., 1995; Schwartz et al., 1998). In addition, the DNA-bending integration host factor (IHF) may participate in hydrogenase promotor activation in *C. necator* (Zimmer et al., 1995; Schwartz et al., 1998). Accordingly, the introduction of this heterologous sigma factor or promoter engineering in *Synechocystis* should be pursued, which was beyond the scope of this study. However, as a proof of principle, we introduced HoxA and HoxJ* into *E. coli* to demonstrate the transcriptional regulation of a *sfgfp* reporter gene fused to P_{SH}. Specific HoxA-dependent GFP fluorescence was detected, likely related to the presence of σ^{54} in *E. coli* (Jones et al., 1994). Our data are consistent with previous findings on functional, HoxA-controlled expression of a reporter gene fused to P_{SH} in *E. coli* (Schwartz et al., 1998). Moreover, as expected, the kinase activity of HoxJ* clearly modulated the HoxA-dependent GFP fluorescence in *E. coli*, leading to a decreased reporter signal. Overall, we now have all genetic elements in hand to eventually assemble a functional H_2 biosensor with optical readout in a cyanobacterium. This could ultimately be used to monitor cyanobacterial H_2 production, e.g., to enable evolutionary or high-throughput screening approaches to improve hydrogenase properties as well as to circumvent existing constraints.

Data availability statement

The original contributions presented in the study are included in the article/Supplementary material, further inquiries can be directed to the corresponding author.

Author contributions

SK designed the study. FO, MAI, and IW constructed the plasmids. FO and IW performed the gene expression and RH activity analyses in *Synechocystis*. FO performed the GFP reporter assays. SL contributed to RH activity determination and experimental expertise for the in-gel assays. LL and OL contributed methodology and know-how on O_2-tolerant hydrogenases. FO and SK wrote the

manuscript with contributions from all co-authors. All authors contributed to the article and approved the submitted version.

Funding

This project was initiated by a grant of the Max-Buchner Foundation to SK (MBFSt-Kennziffer: 3714). We acknowledge the use of the facilities of the Centre for Biocatalysis (MiKat) at the Helmholtz Centre for Environmental Research, which is supported by European Regional Development Funds (EFRE, Europe funds Saxony). We also acknowledge the use of the facilities of H2Saxony. This project (Nr. 100361842) is financed from funds of the European Regional Development Fund (EFRE) and co-financed by means of taxation based on the budget adopted by the representatives of the Landtag of Saxony.

Acknowledgments

The authors thank Jens Appel (University of Kassel, Germany) for the fruitful discussions at the beginning of the project.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

Supplementary material

The Supplementary material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fmicb.2023.1122078/full#supplementary-material

References

Appel, J., Hueren, V., Boehm, M., and Gutekunst, K. (2020). Cyanobacterial *in vivo* solar hydrogen production using a photosystem I–hydrogenase (PsaD-HoxYH) fusion complex. *Nat. Energy* 5, 458–467. doi: 10.1038/s41560-020-0609-6

Ash, P. A., Liu, J., Coutard, N., Heidary, N., Horch, M., Gudim, I., et al. (2015). Electrochemical and infrared spectroscopic studies provide insight into reactions of the NiFe regulatory hydrogenase from *Ralstonia eutropha* with O_2 and CO. *J. Phys. Chem. B* 119, 13807–13815. doi: 10.1021/acs.jpcb.5b04164

Baebprasert, W., Jantaro, S., Khetkorn, W., Lindblad, P., and Incharoensakdi, A. (2011). Increased H_2 production in the cyanobacterium *Synechocystis* sp. strain PCC 6803 by redirecting the electron supply via genetic engineering of the nitrate assimilation pathway. *Metab. Eng.* 13, 610–616. doi: 10.1016/j.ymben.2011.07.004

Barahona, E., Jiménez-Vicente, E., and Rubio, L. M. (2016). Hydrogen overproducing nitrogenases obtained by random mutagenesis and high-throughput screening. *Sci. Rep.* 6:38291. doi: 10.1038/srep38291

Behle, A., Saake, P., Germann, A. T., Dienst, D., and Axmann, I. M. (2020). Comparative dose-response analysis of inducible promoters in cyanobacteria. *ACS Synth. Biol.* 9, 843–855. doi: 10.1021/acssynbio.9b00505

Bernhard, M., Buhrke, T., Bleijlevens, B., de Lacey, A. L., Fernandez, V. M., Albracht, S. P., et al. (2001). The $H_{(2)}$ sensor of *Ralstonia eutropha*. Biochemical characteristics, spectroscopic properties, and its interaction with a histidine protein kinase. *J. Biol. Chem.* 276, 15592–15597. doi: 10.1074/jbc.M009802200

Berto, P., D'Adamo, S., Bergantino, E., Vallese, F., Giacometti, G. M., and Costantini, P. (2011). The cyanobacterium *Synechocystis* sp. PCC 6803 is able to express an active FeFe-hydrogenase without additional maturation proteins. *Biochem. Biophys. Res. Commun.* 405, 678–683. doi: 10.1016/j.bbrc.2011.01.095

Black, L. K., Fu, C., and Maier, R. J. (1994). Sequences and characterization of *hupU* and *hupV* genes of *Bradyrhizobium japonicum* encoding a possible nickel-sensing complex involved in hydrogenase expression. *J. Bacteriol.* 176, 7102–7106. doi: 10.1128/jb.176.22.7102-7106.1994

Bolay, P., Schlüter, S., Grimm, S., Riediger, M., Hess, W. R., and Klähn, S. (2022). The transcriptional regulator RbcR controls ribulose-1,5-bisphosphate carboxylase/oxygenase (RuBisCO) genes in the cyanobacterium *Synechocystis* sp. PCC 6803. *New Phytol.* 235, 432–445. doi: 10.1111/nph.18139

Brandenburg, F., Theodosiou, E., Bertelmann, C., Grund, M., Klähn, S., Schmid, A., et al. (2021). Trans-4-hydroxy-L-proline production by the cyanobacterium *Synechocystis* sp. PCC 6803. *Metab. Eng. Commun.* 12:e00155. doi: 10.1016/j.mec.2020.e00155

Bühler, K., Bühler, B., Klähn, S., Krömer, J. O., Dusny, C., and Schmid, A. (2021). "11 biocatalytic production of white hydrogen from water using cyanobacteria" in *Photosynthesis: Biotechnological applications with microalgae*. ed. M. Rögner (Berlin, Boston: DE GRUYTER), 279–306.

Buhrke, T., Bleijlevens, B., Albracht, S. P., and Friedrich, B. (2001). Involvement of *hyp* gene products in maturation of the $H_{(2)}$-sensing NiFe hydrogenase of *Ralstonia eutropha*. *J. Bacteriol.* 183, 7087–7093.2001, doi: 10.1128/JB.183.24.7087-7093.2001

Buhrke, T., Lenz, O., Krauss, N., and Friedrich, B. (2005). Oxygen tolerance of the H2-sensing NiFe hydrogenase from *Ralstonia eutropha* H16 is based on limited access of oxygen to the active site. *J. Biol. Chem.* 280, 23791–23796. doi: 10.1074/jbc.M503260200

Buhrke, T., Lenz, O., Porthun, A., and Friedrich, B. (2004). The H_2-sensing complex of *Ralstonia eutropha*: interaction between a regulatory NiFe hydrogenase and a histidine protein kinase. *Mol. Microbiol.* 51, 1677–1689. doi: 10.1111/j.1365-2958.2003.03933.x

Bürstel, I., Siebert, E., Frielingsdorf, S., Zebger, I., Friedrich, B., and Lenz, O. (2016). CO synthesized from the central one-carbon pool as source for the iron carbonyl in O_2-tolerant NiFe-hydrogenase. *Proc. Natl. Acad. Sci. U. S. A.* 113, 14722–14726. doi: 10.1073/pnas.1614656113

Dischert, W., Vignais, P. M., and Colbeau, A. (1999). The synthesis of *Rhodobacter capsulatus* HupSL hydrogenase is regulated by the two-component HupT/HupR system. *Mol. Microbiol.* 34, 995–1006. doi: 10.1046/j.1365-2958.1999.01660.x

Elsen, S., Duché, O., and Colbeau, A. (2003). Interaction between the H_2 sensor HupUV and the histidine kinase HupT controls HupSL hydrogenase synthesis in *Rhodobacter capsulatus*. *J. Bacteriol.* 185, 7111–7119. doi: 10.1128/JB.185.24.7111-7119.2003

Englund, E., Liang, F., and Lindberg, P. (2016). Evaluation of promoters and ribosome binding sites for biotechnological applications in the unicellular cyanobacterium *Synechocystis* sp. PCC 6803. *Sci. Rep.* 6:36640. doi: 10.1038/srep36640

Fan, Q., Caserta, G., Lorent, C., Zebger, I., Neubauer, P., Lenz, O., et al. (2022). High-yield production of catalytically active regulatory NiFe-hydrogenase from *Cupriavidus necator* in *Escherichia coli*. *Front. Microbiol.* 13:894375. doi: 10.3389/fmicb.2022.894375

Fernandez-López, R., Ruiz, R., La Cruz, F.De, and Moncalián, G. (2015). Transcription factor-based biosensors enlightened by the analyte. *Front. Microbiol.* 6,:648. doi: 10.3389/fmicb.2015.00648

Gebler, A., Burgdorf, T., de Lacey, A. L., Rüdiger, O., Martinez-Arias, A., Lenz, O., et al. (2007). Impact of alterations near the NiFe active site on the function of the $H_{(2)}$ sensor from *Ralstonia eutropha*. *FEBS J.* 274, 74–85. doi: 10.1111/j.1742-4658.2006.05565.x

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

Greening, C., Biswas, A., Carere, C. R., Jackson, C. J., Taylor, M. C., Stott, M. B., et al. (2016). Genomic and metagenomic surveys of hydrogenase distribution indicate H2 is a widely utilised energy source for microbial growth and survival. *ISME J.* 10, 761–777. doi: 10.1038/ismej.2015.153

Grote, A., Hiller, K., Scheer, M., Munch, R., Nortemann, B., Hempel, D. C., et al. (2005). JCat: a novel tool to adapt codon usage of a target gene to its potential expression host. *Nucleic Acids Res.* 33, W526–W531. doi: 10.1093/nar/gki376

Heidorn, T., Camsund, D., Huang, H.-H., Lindberg, P., Oliveira, P., Stensjö, K., et al. (2011). Synthetic biology in cyanobacteria engineering and analyzing novel functions. *Meth Enzymol.* 497, 539–579. doi: 10.1016/B978-0-12-385075-1.00024-x

Hoffmann, D., Gutekunst, K., Klissenbauer, M., Schulz-Friedrich, R., and Appel, J. (2006). Mutagenesis of hydrogenase accessory genes of *Synechocystis* sp. PCC 6803. Additional homologues of *hypA* and *hypB* are not active in hydrogenase maturation. *FEBS J.* 273, 4516–4527. doi: 10.1111/j.1742-4658.2006.05460.x

Howarth, R. W., and Jacobson, M. Z. (2021). How green is blue hydrogen? *Energy Sci. Eng.* 9, 1676–1687. doi: 10.1002/ese3.956

Huang, H.-H., Camsund, D., Lindblad, P., and Heidorn, T. (2010). Design and characterization of molecular tools for a synthetic biology approach towards developing cyanobacterial biotechnology. *Nucleic Acids Res.* 38, 2577–2593. doi: 10.1093/nar/gkq164

Immethun, C. M., Ng, K. M., DeLorenzo, D. M., Waldron-Feinstein, B., Lee, Y.-C., and Moon, T. S. (2016). Oxygen-responsive genetic circuits constructed in *Synechocystis* sp. PCC 6803. *Biotechnol. Bioeng.* 113, 433–442. doi: 10.1002/bit.25722

Inaba, Y., Morioka, R., Junaid, M., Shiraiwa, Y., and Suzuki, I. (2018). Development of engineered sensor perceiving gaseous toluene signal in the cyanobacterium *Synechocystis* sp. PCC 6803. *J. Appl. Phycol.* 30, 71–78. doi: 10.1007/s10811-017-1277-1

Jones, D. H., Franklin, F. C., and Thomas, C. M. (1994). Molecular analysis of the operon which encodes the RNA polymerase sigma factor sigma 54 of Escherichia coli. *Microbiology* 140, 1035–1043. doi: 10.1099/13500872-140-5-1035

Kleihues, L., Lenz, O., Bernhard, M., Buhrke, T., and Friedrich, B. (2000). The H_2 sensor of *Ralstonia eutropha* is a member of the subclass of regulatory NiFe hydrogenases. *J. Bacteriol.* 182, 2716–2724. doi: 10.1128/JB.182.10.2716-2724.2000

Lacey, R. F., Ye, D., and Ruffing, A. M. (2019). Engineering and characterization of copper and gold sensors in *Escherichia coli* and *Synechococcus* sp. PCC 7002. *Appl. Microbiol. Biotechnol.* 103, 2797–2808. doi: 10.1007/s00253-018-9490-7

Lenz, O., and Friedrich, B. (1998). A novel multicomponent regulatory system mediates H_2 sensing in *Alcaligenes eutrophus*. *Proc. Natl. Acad. Sci. U. S. A.* 95, 12474–12479. doi: 10.1073/pnas.95.21.12474

Lenz, O., Lauterbach, L., Frielingsdorf, S., and Friedrich, B. (2015). "4 oxygen-tolerant hydrogenases and their biotechnological potential" in *Biohydrogen*. ed. M. Rögner (Berlin, München, Boston: DE GRUYTER), 61–69.

Lenz, O., Strack, A., Tran-Betcke, A., and Friedrich, B. (1997). A hydrogen-sensing system in transcriptional regulation of hydrogenase gene expression in *Alcaligenes* species. *J. Bacteriol.* 179, 1655–1663. doi: 10.1128/jb.179.5.1655-1663.1997

Lenz, O., Zebger, I., Hamann, J., Hildebrandt, P., and Friedrich, B. (2007). Carbamoylphosphate serves as the source of $CN^{(-)}$, but not of the intrinsic CO in the active site of the regulatory NiFe-hydrogenase from *Ralstonia eutropha*. *FEBS Lett.* 581, 3322–3326. doi: 10.1016/j.febslet.2007.06.027

Li, T., Jiang, Q., Huang, J., Aitchison, C. M., Huang, F., Yang, M., et al. (2020). Reprogramming bacterial protein organelles as a nanoreactor for hydrogen production. *Nat. Commun.* 11:5448. doi: 10.1038/s41467-020-19280-0

Löscher, S., Gebler, A., Stein, M., Sanganas, O., Buhrke, T., Zebger, I., et al. (2010). Protein-protein complex formation affects the Ni-Fe and Fe-S centers in the H_2-sensing regulatory hydrogenase from *Ralstonia eutropha* H16. *ChemPhysChem* 11, 1297–1306. doi: 10.1002/cphc.200901007

Lubitz, W., Ogata, H., Rüdiger, O., and Reijerse, E. (2014). Hydrogenases. *Chem. Rev.* 114, 4081–4148. doi: 10.1021/cr4005814

Lupacchini, S., Appel, J., Stauder, R., Bolay, P., Klähn, S., Lettau, E., et al. (2021). Rewiring cyanobacterial photosynthesis by the implementation of an oxygen-tolerant hydrogenase. *Metab. Eng.* 68, 199–209. doi: 10.1016/j.ymben.2021.10.006

Mahidhara, G., Burrow, H., Sasikala, C., and Ramana, C. V. (2019). Biological hydrogen production: molecular and electrolytic perspectives. *World J. Microbiol. Biotechnol.* 35:116. doi: 10.1007/s11274-019-2692-z

Martínez-García, E., Goñi-Moreno, A., Bartley, B., McLaughlin, J., Sánchez-Sampedro, L., Pascual del Pozo, H., et al. (2020). SEVA 3.0: an update of the standard European vector architecture for enabling portability of genetic constructs among diverse bacterial hosts. *Nucleic Acids Res.* 48, D1164–D1170. doi: 10.1093/nar/gkz1024

Ni, C., Dinh, C. V., and Prather, K. L. J. (2021). Dynamic control of metabolism. *Annu. Rev. Chem. Biomol. Eng.* 12, 519–541. doi: 10.1146/annurev-chembioeng-091720-125738

Opel, F., Siebert, N. A., Klatt, S., Tüllinghoff, A., Hantke, J. G., Toepel, J., et al. (2022). Generation of synthetic shuttle vectors enabling modular genetic engineering of cyanobacteria. *ACS Synth. Biol.* 11, 1758–1771. doi: 10.1021/acssynbio.1c00605

Patyi, G., Hódi, B., Solymosi, D., Vass, I., and Kós, P. B. (2021). Increased sensitivity of heavy metal bioreporters in transporter deficient *Synechocystis* PCC 6803 mutants. *PLoS One* 16:e0261135. doi: 10.1371/journal.pone.0261135

Pédelacq, J.-D., Cabantous, S., Tran, T., Terwilliger, T. C., and Waldo, G. S. (2006). Engineering and characterization of a superfolder green fluorescent protein. *Nat. Biotechnol.* 24, 79–88. doi: 10.1038/nbt1172

Pierik, A. J., Schmelz, M., Lenz, O., Friedrich, B., and Albracht, S. P. (1998). Characterization of the active site of a hydrogen sensor from *Alcaligenes eutrophus*. *FEBS Lett.* 438, 231–235. doi: 10.1016/S0014-5793(98)01306-4

Ponti, V., Dianzani, M. U., Cheeseman, K., and Slater, T. F. (1978). Studies on the reduction of nitroblue tetrazolium chloride mediated through the action of NADH and phenazine methosulphate. *Chem. Biol. Interact.* 23, 281–291. doi: 10.1016/0009-2797(78)90090-x

Registry of Standard Biological Parts (2003). Available at: http://parts.igem.org (Accessed October 2022).

Riaz-Bradley, A. (2019). Transcription in cyanobacteria: a distinctive machinery and putative mechanisms. *Biochem. Soc. Trans.* 47, 679–689. doi: 10.1042/BST20180508

Schulz, A.-C., Frielingsdorf, S., Pommerening, P., Lauterbach, L., Bistoni, G., Neese, F., et al. (2020). Formyltetrahydrofolate Decarbonylase synthesizes the active site CO ligand of O_2-tolerant NiFe hydrogenase. *J. Am. Chem. Soc.* 142, 1457–1464. doi: 10.1021/jacs.9b11506

Schwartz, E., Gerischer, U., and Friedrich, B. (1998). Transcriptional regulation of *Alcaligenes eutrophus* hydrogenase genes. *J. Bacteriol.* 180, 3197–3204. doi: 10.1128/JB.180.12.3197-3204.1998

Schwartz, E., Henne, A., Cramm, R., Eitinger, T., Friedrich, B., and Gottschalk, G. (2003). Complete nucleotide sequence of pHG1: a *Ralstonia eutropha* H16 megaplasmid encoding key enzymes of H_2-based Lithoautotrophy and Anaerobiosis. *J. Mol. Biol.* 332, 369–383. doi: 10.1016/S0022-2836(03)00894-5

Shafaat, H. S., Rüdiger, O., Ogata, H., and Lubitz, W. (2013). NiFe hydrogenases: a common active site for hydrogen metabolism under diverse conditions. *Biochim. Biophys. Acta* 1827, 986–1002. doi: 10.1016/j.bbabio.2013.01.015

Shcolnick, S., Shaked, Y., and Keren, N. (2007). A role for mrgA, a DPS family protein, in the internal transport of Fe in the cyanobacterium *Synechocystis* sp. PCC 6803. *Biochim. Biophys. Acta* 1767, 814–819. doi: 10.1016/j.bbabio.2006.11.015

Sonntag, C. K., Flachbart, L. K., Maass, C., Vogt, M., and Marienhagen, J. (2020). A unified design allows fine-tuning of biosensor parameters and application across bacterial species. *Metab. Eng. Commun* 11:e00150. doi: 10.1016/j.mec.2020.e00150

Stanier, R. Y., Deruelles, J., Rippka, R., Herdman, M., and Waterbury, J. B. (1979). Generic assignments, strain histories and properties of pure cultures of cyanobacteria. *Microbiology* 111, 1–61. doi: 10.1099/00221287-111-1-1

van Soom, C., de Wilde, P., and Vanderleyden, J. (1997). HoxA is a transcriptional regulator for expression of the *hup* structural genes in free-living *Bradyrhizobium japonicum*. *Mol. Microbiol.* 23, 967–977. doi: 10.1046/j.1365-2958.1997.2781648.x

van Soom, C., Lerouge, I., Vanderleyden, J., Ruiz-Argüeso, T., and Palacios, J. M. (1999). Identification and characterization of *hupT*, a gene involved in negative regulation of hydrogen oxidation in *Bradyrhizobium japonicum*. *J. Bacteriol.* 181, 5085–5089. doi: 10.1128/JB.181.16.5085-5089.1999

Wecker, M. S. A., Beaton, S. E., Chado, R. A., and Ghirardi, M. L. (2017). Development of a *Rhodobacter capsulatus* self-reporting model system for optimizing light-dependent, FeFe-hydrogenase-driven H_2 production. *Biotechnol. Bioeng.* 114, 291–297. doi: 10.1002/bit.26076

Wecker, M. S. A., and Ghirardi, M. L. (2014). High-throughput biosensor discriminates between different algal H_2-photoproducing strains. *Biotechnol. Bioeng.* 111, 1332–1340. doi: 10.1002/bit.25206

Wecker, M. S., Meuser, J. E., Posewitz, M. C., and Ghirardi, M. L. (2011). Design of a new biosensor for algal H_2 production based on the H_2-sensing system of *Rhodobacter capsulatus*. *Int. J. Hydrog. Energy* 36, 11229–11237. doi: 10.1016/j.ijhydene.2011.05.121

Wegelius, A., Khanna, N., Esmieu, C., Barone, G. D., Pinto, F., Tamagnini, P., et al. (2018). Generation of a functional, semisynthetic FeFe-hydrogenase in a photosynthetic microorganism. *Energy Environ. Sci.* 11, 3163–3167. doi: 10.1039/c8ee01975d

Winter, G., Buhrke, T., Jones, A. K., and Friedrich, B. (2004). The role of the active site-coordinating cysteine residues in the maturation of the H_2-sensing NiFe hydrogenase from *Ralstonia eutropha* H16. *Arch. Microbiol.* 182, 138–146. doi: 10.1007/s00203-004-0680-6

Zhang, J., Jensen, M. K., and Keasling, J. D. (2015). Development of biosensors and their application in metabolic engineering. *Curr. Opin. Chem. Biol.* 28, 1–8. doi: 10.1016/j.cbpa.2015.05.013

Zimmer, D., Schwartz, E., Tran-Betcke, A., Gewinner, P., and Friedrich, B. (1995). Temperature tolerance of hydrogenase expression in *Alcaligenes eutrophus* is conferred by a single amino acid exchange in the transcriptional activator HoxA. *J. Bacteriol.* 177, 2373–2380. doi: 10.1128/jb.177.9.2373-2380.1995

frontiers | Frontiers in Microbiology

TYPE Original Research
PUBLISHED 23 March 2023
DOI 10.3389/fmicb.2023.1111979

OPEN ACCESS

EDITED BY
Xuefeng Lu,
Qingdao Institute of Bioenergy and Bioprocess Technology (CAS),
China

REVIEWED BY
Stephan Klähn,
Helmholtz Association of German Research Centers (HZ),
Germany
Lei Chen,
Tianjin University,
China

*CORRESPONDENCE
Satoru Watanabe
✉ s3watana@nodai.ac.jp

SPECIALTY SECTION
This article was submitted to Microbiotechnology, a section of the journal Frontiers in Microbiology

RECEIVED 30 November 2022
ACCEPTED 27 February 2023
PUBLISHED 23 March 2023

CITATION
Sakamaki Y, Maeda K, Nimura-Matsune K, Chibazakura T and Watanabe S (2023) Characterization of a cyanobacterial rep protein with broad-host range and its utilization for expression vectors. *Front. Microbiol.* 14:1111979. doi: 10.3389/fmicb.2023.1111979

Characterization of a cyanobacterial rep protein with broad-host range and its utilization for expression vectors

Yutaka Sakamaki[1], Kaisei Maeda[1,2], Kaori Nimura-Matsune[1], Taku Chibazakura[1] and Satoru Watanabe[1]*

[1]Department of Bioscience, Tokyo University of Agriculture, Tokyo, Japan, [2]Laboratory for Chemistry and Life Science, Institute of Innovative Research, Tokyo Institute of Technology, Yokohama, Japan

Owing to their photosynthetic capabilities, cyanobacteria are regarded as ecologically friendly hosts for production of biomaterials. However, compared to other bacteria, tools for genetic engineering, especially expression vector systems, are limited. In this study, we characterized a Rep protein, exhibiting replication activity in multiple cyanobacteria and established an expression vector using this protein. Our comprehensive screening using a genomic library of *Synechocystis* sp. PCC 6803 revealed that a certain region encoding a Rep-related protein (here named Cyanobacterial Rep protein A2: CyRepA2) exhibits high autonomous replication activity in a heterologous host cyanobacterium, *Synechococcus elongatus* PCC 7942. A reporter assay using GFP showed that the expression vector pYS carrying CyRepA2 can be maintained in not only *S.* 6803 and *S.* 7942, but also *Synechococcus* sp. PCC 7002 and *Anabaena* sp. PCC 7120. In *S.* 7942, GFP expression in the pYS-based system was tightly regulated by IPTG, achieving 10-fold higher levels than in the chromosome-based system. Furthermore, pYS could be used together with the conventional vector pEX, which was constructed from an endogenous plasmid in *S.* 7942. The combination of pYS with other vectors is useful for genetic engineering, such as modifying metabolic pathways, and is expected to improve the performance of cyanobacteria as bioproduction chassis.

KEYWORDS

cyanobacteria, expression vector, broad host range, library screening, next-generation sequencing

1. Introduction

Cyanobacteria are the predominant phototrophs in ocean and freshwater ecosystems, and are among the most widespread phylogenetic clades. Cyanobacteria have oxygen-producing photosynthetic capabilities, meaning that they can produce biomass using solar energy and CO_2, and have recently gained attention for their potential as green cell factories for CO_2-neutral biosynthesis of various products (Knoot et al., 2018; Farrokh et al., 2019). Recently, three cyanobacterial model strains *Synechocystis* sp. PCC 6803 (*S.* 6803), *Synechococcus elongatus* PCC 7942 (*S.* 7942), and *Synechococcus* sp. PCC 7002 (*S.* 7002) have been used in synthetic biology studies for the biosynthesis of multiple products including biofuels (Lan and Liao, 2011; Liu et al., 2011), antioxidants (Shimada et al., 2020), flavors/fragrances (Formighieri and Melis, 2015), and pharmaceuticals (Choi et al., 2016). The

multicellular filamentous cyanobacterium, *Anabaena* sp. PCC 7120 (*A*. 7120), which performs fixation of both nitrogen and carbon, is particularly suitable for the production of nitrogenous substances, and has recently been studied for ammonia production (Higo et al., 2016).

In several model cyanobacteria, *S*. 7942 (Taton et al., 2020), *S*. 6803 (Englund et al., 2015; Chin et al., 2018), and *S*. 7002 (Wang et al., 2019), integration of genetic constructs into chromosomal neutral sites has been used for exogenous gene expression. Due to their polyploidy, chromosomal expression systems in these cyanobacteria are expected to be more effective than those in monoploid organisms, although genetic manipulation is limited to a few cyanobacteria that have natural competence and chromosomal recombination abilities. In addition, genetic engineering relying on integration at neutral sites is time-consuming when performed sequentially, because all chromosomes must be segregated to stably retain the genetically modified constructs in polyploid cyanobacteria.

Self-replicating plasmids are maintained autonomously using replication initiation factors (e.g., Rep proteins) encoded in the plasmids, and harnessing replication machinery of the host. The autonomously replicating sequences consist of an initiation factor and its binding sequence, which serves as the replication origin; however, the binding sequence can also function alone if the host's replication initiation factor functions in *trans*. Most plasmid replicons in *Escherichia coli* cannot be used directly in cyanobacteria, and there is little information on autonomously replicating sequences and available plasmid vectors in cyanobacteria. The broad-host-range plasmid RSF1010, a member of the IncQ plasmid, stably replicates in a wide variety of Gram-negative (Meyer, 2009) and Gram-positive bacteria (Trieu-Cuot et al., 1987), including *E. coli* and several strains of cyanobacteria (Cassier-Chauvat et al., 2021). The RSF1010 vectors are maintained in *S*. 6803 cells but are not suitable for overexpressing genes because their copy number is comparable to that of chromosomes (Jin et al., 2018). Other vectors utilizing cyanobacterial endogenous plasmids have also been developed, including pUH24/pANS (and the derivative vector pUC303) in *S*. 7942 (Kuhlemeier et al., 1983; Van der Plas et al., 1992), pCA2.4, pCB2.4 (pSOMA series; Opel et al., 2022), and pCC5.2 (pSCB; Jin et al., 2018) in *S*. 6803, pAQ (pAQ1-EX1) in *S*. 7002 (Miyasaka et al., 1998), and pDU1 (pRL series) in *Nostoc* sp. PCC 7524 (Wolk et al., 1984). However, these vectors have not been tested for their host range, except for pSOMA, which has been recently demonstrated to be maintained in two cyanobacteria *S*. 6803 and *A*. 7120 (Opel et al., 2022). Further developments in cyanobacterial genetic engineering will require autonomously replicating regions that function in a wide range of species.

In this study, toward the development of broad host-range vectors, we screened autonomously replicating regions and found that a region containing Rep-related protein ORF B (here named Cyanobacterial Rep protein A2: CyRepA2) encoded in the plasmid pCC5.2, derived from *S*. 6803, has replication activity in *S*. 7942. Using this region, we constructed an expression vector, pYS, and tested its expression efficiency, host range, and compatibility with other plasmids. Our results not only contribute to elucidating Rep proteins and their regulatory mechanisms, which are poorly known in cyanobacteria, but also improve the performance of cyanobacteria as bioproduction chassis.

2. Materials and methods

2.1. Cyanobacteria strains and growth condition

The freshwater cyanobacteria *Synechococcus elongatus* PCC 7942 (our laboratory strain *S*. 7942 TUA, which lacks the endogenous small plasmid pHU24/pANS; Watanabe et al., 2012), and *Synechocystis* sp. PCC 6803 (sub-strain *S*. 6803 GT-I; Kanesaki et al., 2012) were used in this study. Both strains were cultured in modified BG-11 medium, which contained double the usual amount of sodium nitrate (final concentration 35.3 mM) and 20 mM HEPES-KOH (pH 8.2), while *Anabaena* sp. PCC 7120 was maintained in standard BG11 (Castenholz, 1988). The marine cyanobacterium *Synechococcus* sp. PCC 7002 was cultured in modified A2 medium containing one-third the usual amount of sodium nitrate (final concentration 17.3 mM; Hasunuma et al., 2019). Cultures were grown photoautotrophically at 30°C, using continuous illumination (50 μmol photons $m^{-2} s^{-1}$) and with 2% CO_2 (v/v) bubbling. When appropriate, spectinomycin (*S*. 7942: final concentration 40 μg/mL, *A*. 7120: 10 μg/mL) or chloramphenicol (*S*. 7942 and *S*. 6803: 7.5 μg/mL, *S*. 7002: 20 μg/mL) were added to media.

2.2. Autonomous replication sequencing

Each DNA fragment was amplified using PrimeSTAR DNA polymerase (TaKaRa, Shiga, Japan) and subcloned into vectors using T4 DNA ligase (TaKaRa) or an In-Fusion HD Cloning Kit (TaKaRa). For the construction of a *S*. 6803 genomic library, the ColE1 origin and the chloramphenicol acetyltransferase gene (*cat*) were PCR-amplified from pUC4K (Amersham, Chicago, IL, United States) and pUC303 (Kuhlemeier et al., 1983) plasmids using the respective primer sets (F1/R2, F3/R4, Supplementary Table S1) and recombined by PCR (primers: F3/R2). After digestion by *Bam*HI and dephosphorylation treatment using shrimp alkaline phosphatase (TaKaRa), the DNA fragment was ligated with 1.5–2.5 kbp fragments of the *Sau*3AI-digested *S*. 6803 genomic DNA. After transformation of *E. coli,* we obtained 2.5×10^4 clones (Library A). To determine the regions responsible for replication in heterologous cyanobacteria, we transformed *S*. 7942 cells with the *S*. 6803 genomic library to yield 335 colonies, which were then pooled (Library B). To obtain plasmids that replicated independently of chromosomal integration, the DNA extracted from library B was introduced into *E. coli* again, and 1.4×10^4 transformants were obtained (Library C). Genomic libraries (Libraries A–C) were subjected to comprehensive sequencing analysis.

The sequencing library was constructed using the NEB Next Ultra DNA Library Prep Kit (NEB, Ipswich, MA, United States) and analyzed using a MiSeq sequencer with a paired-end 150 bp sequence read run with the MiSeq reagent kit v3 (Illumina, San Diego, CA, United States). The reads were trimmed using the CLC Genomics Workbench ver. 9.5.4 (QIAGEN, Venlo, Netherlands) with the following parameters: Phred quality score > 30; ambiguous nucleotides allowed:1; automatic read-through adaptor trimming: yes; removing the terminal 15 nucleotides from the 5′ end and five nucleotides from the 3′end; and removing truncated reads of less than 30 nucleotides in length. To identify the regions included in the library, trimmed reads were mapped to the *S*. 6803 genome (accession numbers, chromosome: AP012276, pSYSM:

AP004310, pSYSX: AP006585, pSYSA: AP004311, pSYSG: AP004312, pCA2.4: CP003270, pCB2.4: CP003271, and pCC5.2: CP003272) using the CLC Genomics Workbench ver. 20.0.1 (QIAGEN) with the following parameters: match score, 1; mismatch cost, 2; indel cost, 3; length fraction, 0.8; similarity fraction, 0.9; and non-specific match handling, ignored. The number of raw read pairs per sample and the ratio of reads mapped on the reference sequences are shown in Supplementary Table S2; Supplementary Figure S1A. Original sequence reads were deposited in the DRA/SRA database with the following accession numbers (Library A: DRR285589, Library B: DRR285594, and Library C: 285596). The accession number for the BioProject was PRJDB11466.

2.3. Phylogenetic analysis

The amino acid sequence of ORF B (accession number: WP_015390179.1) was obtained from the database, and its homologs were retrieved by an NCBI-BLAST search for the top 100 most similar sequences. Among these, we excluded overlapping sequences and obtained 58 homologous sequences. Phylogenetic analysis was performed on these homologs, along with pUH24/pANS Rep in *S.* 7942 and RSF1010 RepC, using ClustalW within MEGA11, with default parameters (Tamura et al., 2021). Evolutionary history was inferred using the neighbor-joining method (Saitou and Nei, 1987). The optimal tree is shown in Figure 1A. The percentage of replicate trees in which the associated taxa clustered together in the bootstrap test (1,000 replicates) is shown next to the branches (Felsenstein, 1985). The tree was drawn to scale, with branch lengths in the same units as those of the evolutionary distances used to infer the phylogenetic tree. The evolutionary distances were computed using the Poisson correction method (Zuckerkandl and Pauling, 1965) and were expressed as number of amino acid substitutions per site. This analysis involved 61 amino acid sequences (ORF B, pUH24/pANS Rep, RSF1010 RepC, and 58 homologs). All ambiguous positions were removed for each sequence pair (pairwise deletion). There were 1,655 positions in the final dataset. Evolutionary analyses were conducted using MEGA11 (Tamura et al., 2021).

2.4. Prediction of 3D structure of CyRepAs

AlphaFold2 was used for 3D structure prediction (AlQuraishi, 2019; Cramer, 2021). The predicted 3D structure of Slr7037 (CyRepA1) was obtained from the AlphaFold Protein Structure Database. ORF B (CyRepA2) was predicted by using ColabFold and the model with the best score was adopted (Mirdita et al., 2022). The predicted 3D structures were colored to domains using UCSF Chimera (University of California San Francisco, San Francisco, CA, United States; Pettersen et al., 2004). Domain predictions were performed using SMART (Letunic et al., 2021) and Pfam software (Mistry et al., 2021). Predicted 3D structure data of CyRepA1 and CyRepA2 are available in Supplementary Data 1, 2.

2.5. Plasmid and strain construction

To construct the expression vector pYS1C-GFP, we used a plasmid isolated from *E. coli* transformant pools of Library B (Supplementary Table S3). DNA fragments containing the *lacI* gene, the *trc* promoter, the *gfp* ^{mut2} gene (Cormack et al., 1996), and the B0011 *lux* operon terminator were PCR-amplified using appropriate primer sets (F5/R6, F7/R8, F9/R10, F11/R12, and F13/R14) and introduced into the plasmid obtained from library screening and containing ColE1, *cat*, and a part of pCC5.2 (1,174–4,674 nt) by In-Fusion cloning (Supplementary Figure S2; Supplementary Data 3).

To enable strict control of gene expression in *S.* 7002, the promoter region of pYS1C-GFP was replaced with the *clac143* promoter, which was optimized for regulated expression in *S.* 7002. Using primers F15/R16 flanking the sequence of the *clac143* promoter, DNA fragments were PCR-amplified from pYS1C-GFP and circularized using In-Fusion cloning. The resulting plasmid was designated pYS4C-GFP (Supplementary Figure S2; Supplementary Data 4). To replace the *cat* gene in pYS1C-GFP and pYS4C-GFP with the spectinomycin-resistance (Sp^R) gene, the plasmid regions and Sp^R gene were PCR-amplified using appropriate primer sets (F17/R18, F19/R20) and combined using In-Fusion cloning. The resulting plasmids were designated pYS1S-GFP and pYS4S-GFP.

For the construction of the integration plasmid pNSG, a DNA fragment of the *gfp* ^{mut2} gene was PCR-amplified with primers F21/R22 and cloned into the *Sal*I and *Hind*III sites of the integration vector pNSE1 (Kato et al., 2008), containing a region homologous to a neutral site in *S.* 7942. To construct a plasmid harboring the pUC303-based replication system, the region containing *repA* and *repB* of pUC303, the p15A origin, a DNA fragment containing Sp^R gene, and the *mScarlet* gene under the control of the *cpcB* promoter were PCR-amplified using appropriate primer sets (F23/R24, F25/R26, F27/R28) and combined using In-Fusion cloning to form a plasmid designated pEX2S-mScarlet.

To transform the cyanobacteria *S.* 7942, *S.* 6803, and *S.* 7002, cells were grown as described above until they reached an OD_{750} of 0.7–1.2 and then collected by centrifugation at 3,000 × *g* for 10 min at 25°C. After resuspension in BG11 to 10-fold concentration, plasmid DNA was added and incubated at 25°C overnight under light-shielded and rotating conditions. The samples were then irradiated with light for 1 h and spread on drug-containing plates. In the case of *A.* 7120 transformation, the plasmids pYS1S-GFP or pYS4S-GFP were transformed into *E. coli* HB101 carrying the plasmid pRL623 (Elhai et al., 1997) and were transferred to *A.* 7120 cells with the help of the conjugative plasmid RP4, according to the triparental mating method (Thiel and Wolk, 1987; Elhai and Wolk, 1988) with minor modifications. *Escherichia coli* cultures carrying the donor plasmid were mixed with the *E. coli* culture carrying RP4 and incubated for 1 h. The *A.* 7120 culture prepared as described above was added to the *E. coli* mixture and incubated overnight on a BG11 plate containing 5% LB medium. The mixture was harvested by adding BG11 medium to the plate and spread on plates containing spectinomycin.

The structures of the pYS plasmids introduced into cyanobacteria were confirmed by the following procedure: DNA extracted from cyanobacteria harboring pYS plasmids was used to transform *E. coli* JM109. After selection in LB medium containing chloramphenicol or spectinomycin, plasmid DNA was extracted by using the FavorPrep™ Plasmid DNA Extraction Mini Kit (FAVORGEN, Ping Tung, Taiwan). Plasmid DNA samples were digested with *Nde*I and *Pst*I for pYS1C-GFP and pYS4C-GFP, and *Sal*I and *Nhe*I for pEX2S-mScarlet. Each digested DNA sample (80 ng) was subjected to 1% agarose gel electrophoresis.

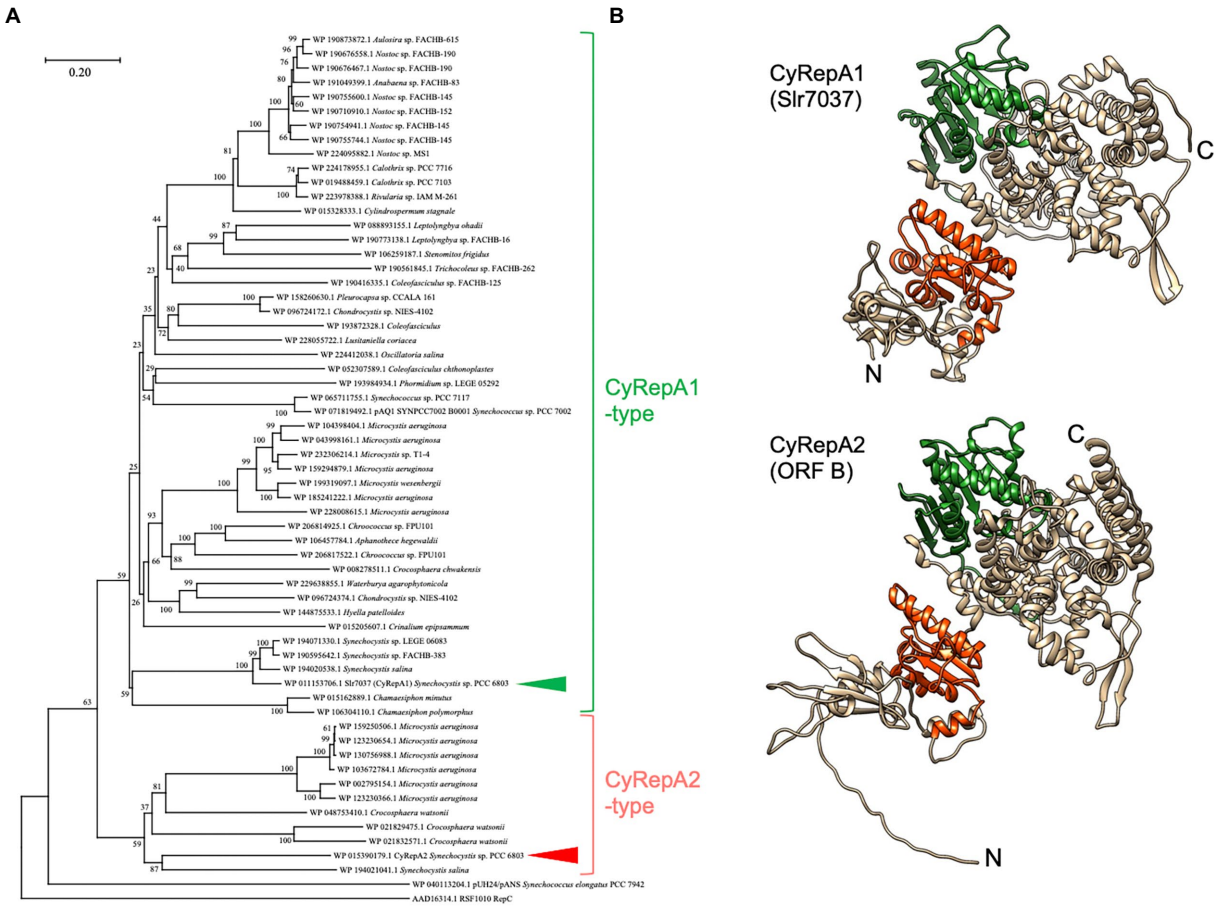

FIGURE 1
Phylogenetic and structural analysis of Cyanobacterial Rep protein A (CyRepA1: Slr7037; CyRepA2: ORF B). **(A)** Phylogenetic tree of CyRepA. CyRepA2 (ORF B) and its 58 homologs were used for phylogenetic analysis along with pUH24/pANS Rep and RSF1010 RepC used as outgroups. Proteins used for analysis were shown by their accession numbers in database with organism names. The tree is drawn to scale, with branch lengths in the same units as those of the evolutionary distances used to infer the phylogenetic tree. CyRepA1 (Slr7037) and CyRepA2 are indicated by arrow heads. **(B)** 3D structures of CyRepA1 and CyRepA2 predicted using AlphaFold2. The 3D structures are color-coded as follows based on the results of domain predictions by SMART and Pfam. DUF3854 domain (orange), DEXDc domain (green).

2.6. SDS-PAGE and western blotting analysis

The protein extracts from cyanobacteria cells were prepared according to the method described by Nimura et al. (2001) and were separated by SDS-PAGE on 12% polyacrylamide gels at 150 V for 90–100 min in electrophoresis buffer (25 mM Tris, 192 mM glycine, and 0.1% SDS). The gels were blotted onto 0.2 μm PVDF membranes (Bio-Rad, Hercules, CA, United States) at 2.5 A for 10 min using a Trans-Blot Turbo Transfer System (Bio-Rad). Membranes were blocked for 30 min with 5% skimmed milk in TBS-T [50 mM Tris–HCl (pH 7.5), 150 mM NaCl, 0.05% Tween 20], washed, and soaked in TBS-t for 2 h at room temperature with anti-GFP (MBL, Tokyo, Japan), anti-RpoD1 rabbit antiserum (Seki et al., 2007), or anti-RbcL (Agrisera, Vännäs, Sweden) as primary antibodies. After washing, the membranes were soaked in TBS-t for 30 min at room temperature with HRP-conjugated anti-mouse (GE Healthcare, Chicago, IL, United States) or anti-rabbit antibody (GE Healthcare) as secondary antibodies. Chemiluminescence was detected with the LumiGLO Chemiluminescent Substrate KPL using ChemiDoc XRS Plus (Bio-Rad).

2.7. Flow cytometry

Cells were analyzed by GFP fluorescence-activated cell sorting (FACS) using a BD Accuri™ C6 Flow cytometer (BD Biosciences, San Jose, CA, United States) and BD CFlow software (BD Biosciences), as described by Watanabe et al. (2012). Cyanobacterial cells showing chlorophyll fluorescence were sorted using the FL3 channel and GFP fluorescence was measured using the FL1 channel.

2.8. Microscopy

Fluorescent images were obtained using BX53 microscope (OLYMPUS, Tokyo, Japan) at ×40 or 100 magnification with a DP71 digital camera (OLYMPUS) and DP Controller software ver. 3. 3. 1. 292 (OLYMPUS).

3. Results

3.1. Library screening of sequences responsible for autonomous replication activity in *Synechococcus elongatus* PCC 7942

To gain new insights into autonomously replicating regions in cyanobacteria, we established a new method, Autonomous Replication sequencing (AR-seq), which combines library screening and sequencing. As a source organism for screening, we selected *S.* 6803 (library A), which contains seven plasmids in addition to the chromosome, and at least seven replicons that can coexist. *S.* 7942 was selected as the host cyanobacteria for transformation with the plasmid-based *S.* 6803 genomic library for three reasons: avoiding chromosome-library recombination, having extremely high transforming ability, and containing only one large plasmid pANL. A genomic library of 2.5×10^4 clones (library A) was constructed by ligating 1.5–2.5 kbp fragments of *S.* 6803 genomic DNA with the ColE1 origin and *cat* gene, and this was introduced into *S.* 7942. As a result, more than 300 transformants were obtained. DNA extracted from *S.* 7942 transformants was pooled as library B, which contained the autonomously replicating sequences in addition to *S.* 7942 genomic DNA. To isolate plasmids that replicated independently of chromosomal integration in *S.* 7942 cells, library B was again introduced into *E. coli*, and only the plasmid DNA was pooled as library C. At this stage, 14 *E. coli* colonies were picked from the transformants of library B, and the inserts were sequenced by Sanger sequencing. The results showed that all colonies contained the region surrounding ORF B of pCC5.2, a plasmid found in *S.* 6803 (Supplementary Table S3).

Next, a comprehensive sequencing analysis was performed to reveal the genomic regions in the libraries. The results showed that pCC5.2 occupied a large proportion of the library in spite of its relatively short DNA length in the *S.* 6803 genomic library A (Supplementary Figure S1A; Supplementary Table S2), suggesting a high copy number of pCC5.2 in *S.* 6803 cells, as reported previously (Jin et al., 2018). The fact that only a few sites were recognized by *Sau*3AI in pCA2.4 and pCB2.3 may have contributed to their relatively small proportion in the library A. Compared to the library A, the read population of pCC5.2 significantly increased in library B, and 99.8% of the reads were occupied by pCC5.2 in library C (Supplementary Figure S1A; Supplementary Table S2). We mapped the sequence reads to pCC5.2 to identify the region required for replication, and observed that the region containing ORF B was concentrated in the sequence reads from libraries B and C (Supplementary Figure S1B). ORF B and the flanking region of pCC5.2 (3,242 nt) have been reported to function as replicons to support plasmid replication in *S.* 6803 (Xu and McFadden, 1997; Jin et al., 2018). Our study suggests that this region functions as a replicon that can support plasmid replication in the heterologous cyanobacterium, *S.* 7942.

3.2. Phylogenetic analysis of ORF B in pCC5.2 (CyRepA2)

To characterize ORF B in pCC5.2, we performed a sequence-based analysis of the conserved regions and structure of ORF B. Analysis of the SMART and Pfam databases revealed that ORF B has a DUF3854 domain in its N-terminal region. BLAST results revealed that DUF3854-containing proteins are widely conserved among cyanobacteria, not only *Synechococcales,* which contain *S.* 6803 and *S.* 7942, but also *Oscillatoriophycideae*, *Nostocaceae*, *Pleurocapsales*, *Pseudanabaenales*, and *Chroococcidiopsidales* (Figure 1A). The DUF3854-containing proteins formed a distinctly different group from pUH/pANS Rep and RSF1010 RepC and were classified into two main groups. The larger group contained Slr7037 and SYNPCC7002_B0001 proteins encoded in pSYSA (*S.* 6803) and pAQ1 (*S.* 7002), which function as RepA homologs in each organism (Miyasaka et al., 1998; Kaltenbrunner et al., 2022). ORF B belongs to a smaller group, containing proteins in *Microcystis*, *Crocosphaera*, and *Synechocystis*. Although Slr7037 and ORF B are classified into different clades, the predicted structures of these proteins are very similar, with an N-terminal DUF3854 domain and a C-terminal DEXDc domain connected by alpha-helices (Figure 1B). DEXDc domains have been found in proteins with DNA helicase activity, and the function of the DUF3854 domain has been suggested to be related to the Toprim domain conserved in DNA primases (Honda et al., 2006; Jin et al., 2018), although its function remains unknown. Since the group containing Slr7037 is more common in cyanobacteria than the group containing ORF B, we designated Slr7037 as Cyanobacterial Rep protein A (CyRepA1), named ORF B as CyRepA2, and conducted further analysis.

3.3. Construction of autonomously replicating plasmid pYS1 in *Synechococcus elongatus* PCC 7942

To evaluate the replication activity of CyRepA2 in *S.* 7942, the expression vector pYS1C-GFP (Supplementary Figure S2) was constructed based on a plasmid containing the minimal region responsible for autonomous replication (1,174–4,683 nt in pCC5.2), ColE1 (replicon for *E. coli*), and the *cat* gene. pYS1C-GFP possesses *gfp* gene under the control of the *trc* promoter and the *lacI* gene, which functions in *S.* 7942, and thus can express GFP in an IPTG-dependent manner. Plasmid pYS1C-GFP was introduced into *S.* 7942 cell and the transformants were used for following analysis. After 48 h of pre-incubation, the cells were cultivated for additional 24 h in presence or absence of 1 mM IPTG and used for analysis. Fluorescence microscopy revealed IPTG-dependent GFP fluorescence in *S.* 7942 cells carrying pYS1C-GFP (Figure 2A). To confirm the structure of pYS1C-GFP in *S.* 7942 cells, DNA was prepared from cells harboring pYS1C-GFP and introduced into *E. coli*. Plasmids were extracted from the resulting *E. coli* transformants and compared with the initial pYS1C-GFP plasmid using restriction enzyme digestion (Figure 3). The results showed that both plasmids were identical, suggesting that pYS1C-GFP was maintained as a plasmid in *S.* 7942 cells.

To compare the utility of pYS1C-GFP with our previous expression system integrated into the chromosomal neutral site, we constructed an integration plasmid pNSG carrying *gfp* gene. The NS-GFP strain, which expresses GFP from a chromosomal neutral site in an IPTG-dependent manner, was obtained by the transformation of *S.* 7942 with pNSG and GFP expression levels were compared by fluorescence microscopy, flow cytometry, and western blot analysis. Microscopy showed that GFP fluorescence of cells carrying pYS1C-GFP was significantly stronger than that of NS-GFP cells (Figure 2A), although some cells without GFP fluorescence were observed. The expression levels of GFP were quantified and compared by flow cytometry and western blotting analysis, showing that its expression in cells carrying pYS1C-GFP was approximately 10-fold

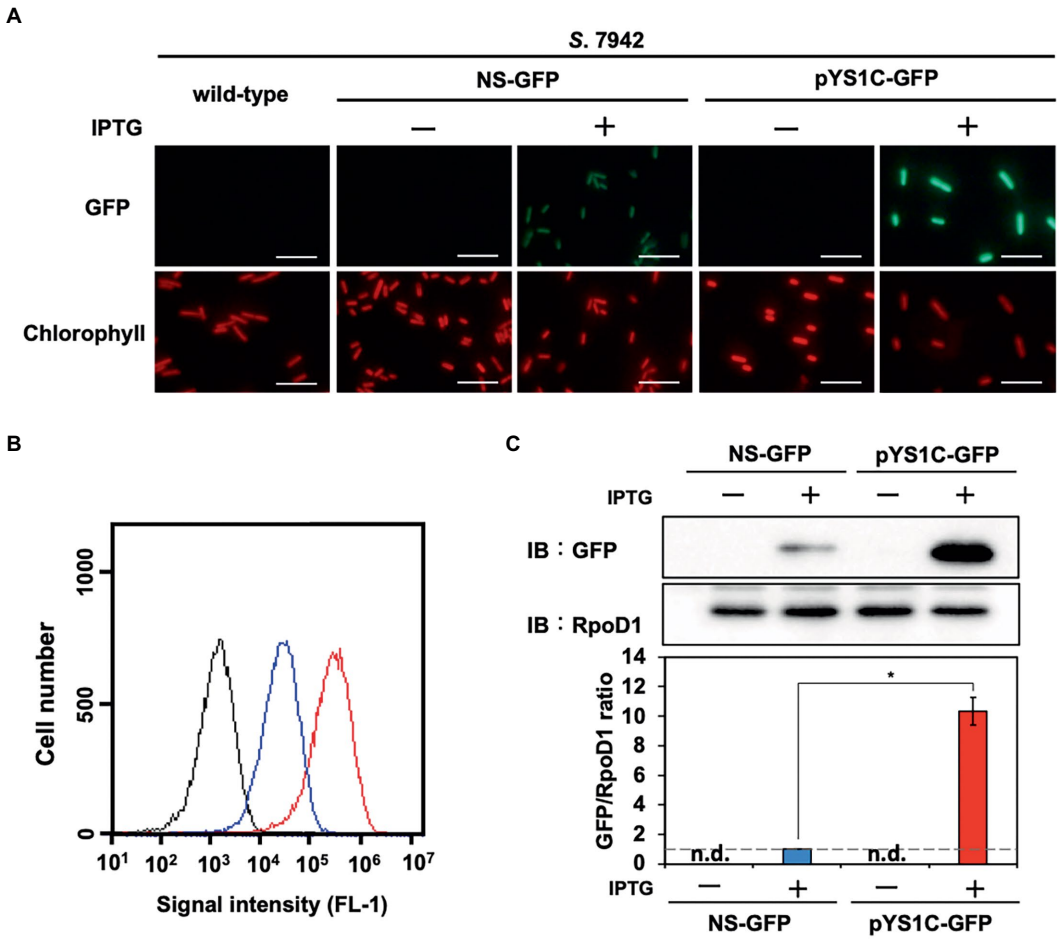

FIGURE 2
Comparison of expression performance between chromosome and pYS1 plasmid by the GFP reporter system in *S.* 7942. The GFP expression levels in the pYS1C-GFP transformant were compared to *Synechococcus elongatus* PCC 7942 (*S.* 7942) wild-type and NS-GFP strains expressing GFP at the chromosomal neutral site. GFP expression was induced by the addition of 1mM IPTG (final concentration). **(A)** Fluorescence microscopy images. The GFP and chlorophyll fluorescence images are shown. White bar, 10μm. **(B)** FACS analysis of the GFP fluorescence. Signal intensity of FL-1 indicating GFP fluorescence in the wild-type (black), NS-GFP (blue), and pYS1-GFP (red) strains are shown. **(C)** Western blot analysis. The protein extracts obtained from wild-type, NS-GFP, and pYS1C-GFP strains were subjected to SDS-PAGE and analyzed by western blotting using antibodies against GFP. RpoD1 was used as an internal control. The signal intensities of GFP were normalized to those of RpoD1, and the ratio the GFP signal to NS-GFP was set to 1. Bars represent the mean±SEM (n=3; p<0.05).

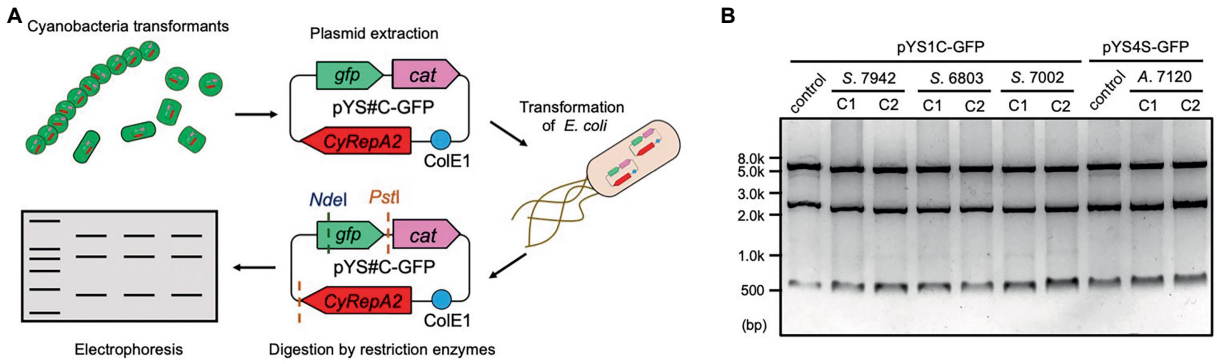

FIGURE 3
Plasmid structure of pYS in cyanobacteria cells. **(A)** Scheme of the analysis of plasmid structure. To determine whether plasmids are maintained in cyanobacterial cells in a circular structure, DNA was extracted from cyanobacteria transformants (*S.* 7942, *S.* 6803, *S.* 7002, and *A.* 7120) carrying pYS1C-GFP or pYS4S-GFP and introduced into *E. coli*. After the plasmid was extracted from the *E. coli* cells, plasmids were digested with restriction enzymes, and compared to the plasmids before the transformation of cyanobacteria. **(B)** Electrophoresis image of plasmids digested by the restriction enzymes. The plasmids before transformation were used as controls. The results of two independent clones are shown as C1 and C2.

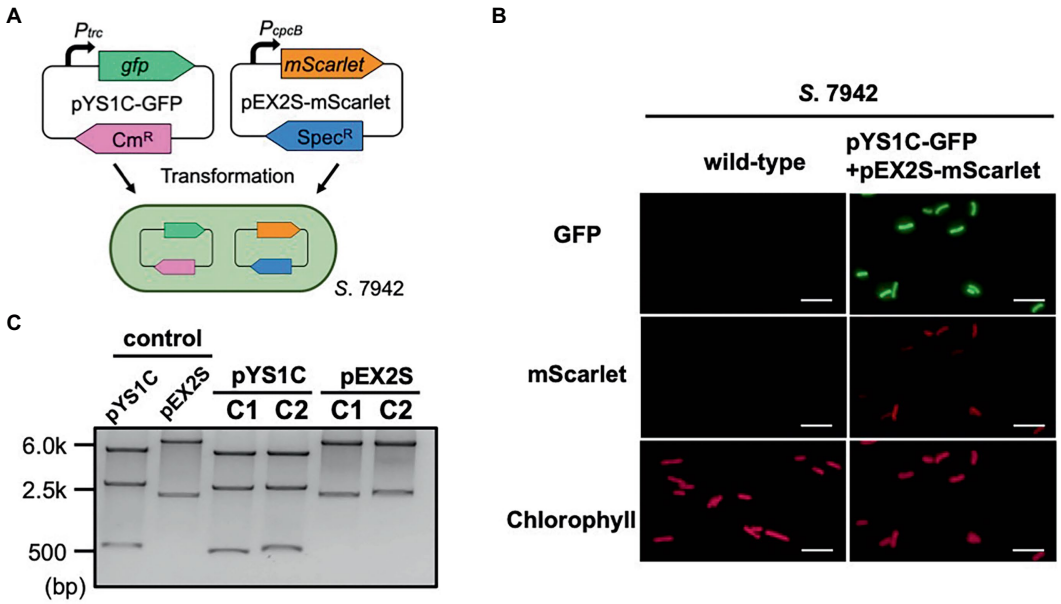

FIGURE 4
Compatibility of pYS and endogenous plasmid-derived vectors in *S.* 7942. **(A)** Compatibility test in *S.* 7942. pEX2S-mScarlet containing spectinomycin resistance gene (Sp^R) with pHU24/pANS Rep was introduced into *S.* 7942 cell carrying pYS1C-GFP and subjected to microscopy and DNA extraction. **(B)** Fluorescence microscopy. GFP expression was induced in pYS1C-GFP by the addition of 1mM IPTG (final concentration) to the culture. Images of GFP, mScarlet, and chlorophyll fluorescence are shown. White bar: 10μm. **(C)** Structure of the plasmids before and after the transformation of *S.* 7942. Plasmid DNA was extracted from *S.* 7942 transformant and transformed into *E. coli*. After extraction from *E. coli*, plasmids were digested with restriction enzymes, and compared with the plasmid before the transformation of *S.* 7942 (control). *Nde*I and *Pst*I (for pYS1C-GFP) and *Sal*I and *Nhe*I (for pEX2S-mScarlet) were used for the digestion. The plasmids before transformation were used as controls. The results of two independent clones are shown as C1 and C2.

higher than that in NS-GFP cells and was solely dependent on IPTG (Figures 2B,C). These results indicated that the expression system using pYS1 harboring CyRepA2 is more suitable for overexpression than that integrated into the chromosomes in *S.* 7942.

The stability of pYS1C-GFP and the expression level of GFP in *S.* 6803 were also tested using the same approaches as in *S.* 7942, because the use of pCC5.2 as an expression vector has been reported in *S.* 6803 cells (Jin et al., 2018). Consistent with a previous study, we confirmed that pYS1C-GFP was maintained autonomously as a plasmid within *S.* 6803 cells (Figure 3). Although a 5-fold increase in GFP expression was observed with IPTG addition in *S.* 6803, a leaky expression was also detected in the absence of IPTG, in contrast to *S.* 7942 (Supplementary Figure S3). In addition, PCR analysis indicated that the *S.* 6803 pYS1C-GFP transformants contained endogenous pCC5.2 as well, suggesting the instability of pYS1C-GFP (Supplementary Figure S4). Using a more tightly regulated promoter for pYS1 in *S.* 6803 would achieve a more precise expression and stable maintenance of this plasmid.

3.4. Compatibility of pYS and endogenous plasmid-derived vector in *Synechococcus elongatus* PCC 7942

Phylogenetic analysis revealed that the sequence of CyRepA2 differs from that of the RepA protein encoded in the plasmid pUH24/pANS of *S.* 7942 (Figure 1A). This led us to assume that pYS can be maintained along with the pHU24/pANS-derived vector in *S.* 7942 cells. To investigate the compatibility of Rep proteins, we constructed the plasmid pEX2S-mScarlet expressing the orange fluorescent protein mScarlet with the spectinomycin resistance marker gene, based on the plasmid pUC303, which was derived from pHU24/pANS. As a result of the transformation of *S.* 7942 cells carrying pYS1C-GFP with pEX2S-mScarlet, we obtained colonies that showed resistance to both chloramphenicol and spectinomycin (Figure 4A). Fluorescence microscopy revealed both GFP and mScarlet fluorescence in *S.* 7942 transformants (Figure 4B), indicating that this strain harbors pEX2S-mScarlet with pYS1C-GFP. In addition, the GFP expression levels of cells carrying these two plasmids were comparable to those of cells harboring only pYS1C-GFP (Supplementary Figure S5). To determine the plasmid structure in *S.* 7942 cells, DNA was extracted, introduced into *E. coli* and selected using chloramphenicol or spectinomycin. Plasmid DNA extracted from the *E. coli* transformants was digested with restriction enzymes. The results showed that the two plasmids retained the same structure as before transformation, confirming that they were independently maintained in *S.*7942 cells (Figure 4C).

3.5. Development of expression vectors functioning in *Synechococcus* sp. PCC 7002

A DUF3854-containing protein was predicted to be encoded in the pAQ1 plasmid of the model marine cyanobacterium *S.* 7002

(SYNPCC7002_B0001), which is closely related to CyRepA1 rather than CyRepA2 (Figure 1A). Thus, we tested the availability of pYS1 in *S.* 7002 using a similar procedure for both *S.* 7942 and *S.* 6803. Similar to the previous two cyanobacteria, significant GFP expression and plasmid maintenance were observed in *S.* 7002 (Figures 3, 5A–D). To test the compatibility between pYS plasmid and endogenous pAQ1 in *S.* 7002 cells, we performed PCR analysis using specific primer sets for the amplification of pYS1C-GFP and pAQ1. The results showed that pAQ1 was stably retained in the pYS1C-GFP transformant (Supplementary Figure S4), indicating that the replication system of the CyRepA2-type can be coexist with that of the CyRepA1-type despite the similarity of their amino-acid sequences. In contrast to the cases of *S.* 7942 and *S.* 6803, there was no difference between the expression (both fluorescence and amount) of GFP proteins in the presence and absence of IPTG in *S.* 7002 (Figures 5A–D), indicating that repression of the *trc* promoter by LacI does not work at all in *S.* 7002, as reported previously (Markley et al., 2015). We noted a broad distribution of the weaker fluorescence signal in the FACS profile of *S.* 7002 (Figure 5B), suggesting that there were differences in the GFP expression levels varies in each cell, which is consistent with the GFP fluorescence observed by microscopy.

To improve the induction system of pYS1C-GFP in *S.* 7002 cells, the *trc* promoter was replaced with the *clac143* promoter, which enables IPTG-dependent regulation in *S.* 7002 (Markley et al., 2015). The resulting plasmid, designated pYS4C-GFP, was introduced into *S.* 7002 and GFP expression was tested in the presence or absence of IPTG. SDS-PAGE analysis demonstrated that the expression level of GFP could be controlled by IPTG (Figures 5E–H). Comparison of the GFP fluorescence intensity of pYS1C-GFP and pYS4C-GFP strains in the presence of IPTG by flow cytometry showed that the fluorescence of the cells carrying pYS1C-GFP was more than one order of magnitude higher than that of cells carrying pYS4C-GFP, indicating that pYS1 can be used as an overexpression system and pYS4 can be used in the situation when strict control of expression is required (Figures 5B,F).

3.6. Utilization of pYS vector In *Anabaena* sp. PCC 7120

To further test the utility of the pYS plasmid, we introduced it *via* conjugation to filamentous nitrogen-fixing cyanobacterium *A.* 7120, which belongs to *Nostocaceae,* a phylogenetic group distinct from *Synechococcales* including *S.* 6803, *S.* 7942, and *S.* 7002. For the conjugative transport of the pYS plasmid to *A.* 7120, which requires a helper plasmid pRL carrying the *cat* marker gene, we replaced the *cat* gene in pYS1C- and pYS4C-GFP with the spectinomycin resistance marker gene and named the resulting plasmids pYS1S-GFP and pYS4S-GFP, respectively. We successfully isolated 10 clones of *A.* 7120 transconjugants carrying pYS4S-GFP, suggesting that this plasmid could be transported to *A.* 7120 cells by conjugation at a low frequency. On the other hand, a transconjugant carrying pYS1S-GFP was not obtained. This may be due to the overexpression of GFP by the *trc* promoter, as in *S.* 7002, which could be too strong and cytotoxic in *A.* 7120 cells. In the pYS4S-GFP transconjugants, GFP fluorescence was observed in every filamentous cell and the structure of pYS4S-GFP was identical to that of the plasmid before transconjugation (Figures 3, 6A), suggesting that the pYS4-derived vector was maintained stably

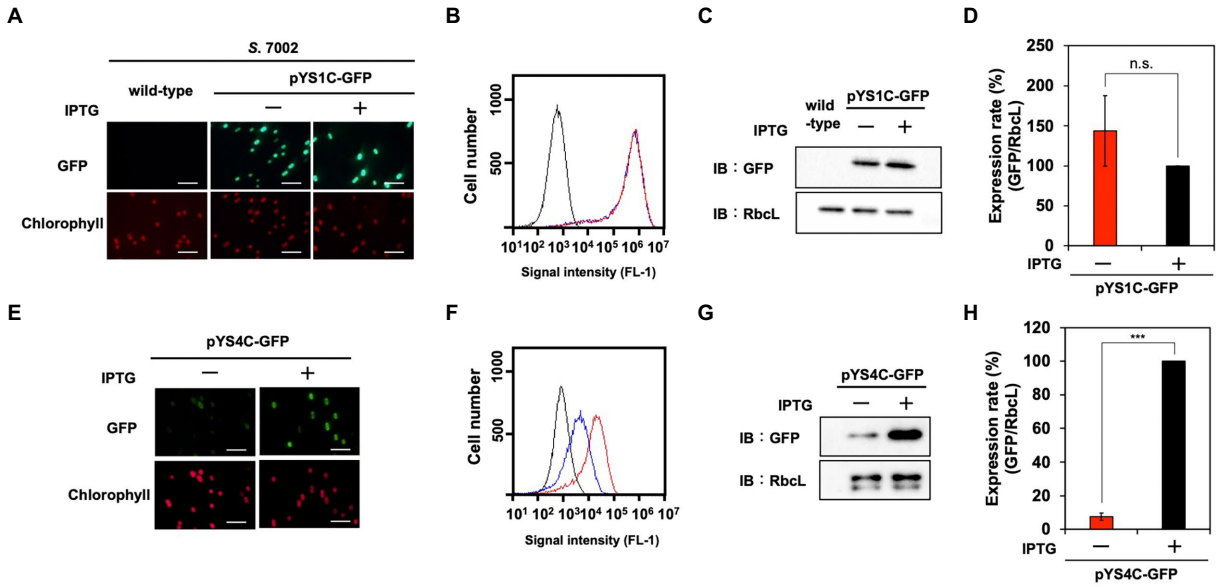

FIGURE 5
Utilization of expression plasmid pYS1 and pYS4 in *S.* 7002. The GFP expression levels in pYS1C-GFP **(A–D)** and pYS4C-GFP transformants **(E–H)** were analyzed with (+) and without (–) 1mM IPTG. **(A,E)** Fluorescence microscopy images. The GFP and chlorophyll images are shown. White bar: 10μm **(B,F)** FACS analysis of GFP fluorescence. Signal intensity of FL-1 indicating GFP fluorescence in *S.* 7002 wild-type (black), pYS- or pYS4-GFP in the presence (red) and absence (blue) of IPTG are shown. **(C,G)** Western blotting analysis. The protein extracts obtained from *S.* 7002 cells were subjected to SDS-PAGE and analyzed by western blotting using antibodies against GFP. RbcL was used as an internal control. **(D,H)** Comparison of GFP signals. The signal intensity of GFP obtained from western blot analysis was normalized to that of RbcL, and the ratio of GFP signal in the presence of IPTG was set to 100. Bars represent the mean±SEM (n=3; ***p<0.001).

in *A.* 7120 cells; however, in contrast to *S.* 7002, there was no difference between GFP expression in the presence and absence of IPTG (Figures 6B,C).

4. Discussion

Cyanobacteria, which grow by absorbing CO_2 through photosynthesis, are promising hosts for carbon-neutral material production. However, even with model cyanobacteria, such as *S.* 7942 and *S.* 6803, only a limited number of vectors are available for genetic engineering. Library screening of autonomously replicating regions revealed that the CyRepA2-containing region derived from *S.* 6803 functions in *S.* 7942 and can be used in overexpression vector that function in a wide range of cyanobacterial species.

Our AR-seq analysis using the *S.* 6803 genomic library revealed that the open reading frame (ORF) of CyRepA2 and its upstream region are required for autonomous replication in *S.* 7942 (Supplementary Figure S1B; Supplementary Table S3). This is the first report showing that the CyRepA2 region has autonomous replication activity in heterologous cyanobacteria as well as in its original host, *S.* 6803 (Jin et al., 2018). Although only one autonomous replication region was obtained from this screening, it is possible to identify new Rep and autonomously replicating regions by changing the genomic library and host organisms used for AR-seq. Given its versatility, AR-seq could be used to identify autonomously replicating initiation regions not only in cyanobacteria but also in a variety of other organisms.

It has been reported that pCC5.2 exhibits rolling-circle replication (RCR; Xu and McFadden, 1997). Many RCR plasmids have been shown to replicate in species, genera, or even phyla other than those from which they were isolated. The simplicity of RCR initiation, with only the plasmid-encoded Rep protein participating in origin recognition and priming of leading strand synthesis, is thought to underlie the unusual promiscuity of plasmids (del Solar et al., 2014). Consistent with these observations, the pYS plasmid constructed in this study based on pCC5.2 also exhibited a wide host range, similar to the RCR-employing pSOMA plasmids derived from pCA2.4 and pCB2.4 (Opel et al., 2022). RCR plasmids appear to be suitable for expression because of their small size, high copy number, and the few specific factors involved in replication. Especially in *S.* 7942, GFP expression was 10-fold higher than that in chromosome-based expression systems, which can be tightly controlled by IPTG. These results strongly suggest that the pYS vectors constructed in this study are exceptional genetic engineering tools for potential use at an industrial level in cyanobacteria.

Other applications of RCR plasmids include their use in combination with other plasmids. In this study, we showed that pYS derived from pCC5.2 can be harbored simultaneously with pEX2, derived from an endogenous plasmid in *S.* 7942. Furthermore, we demonstrated that pAQ1 can be maintained along with pYS1C-GFP in *S.* 7002 cells (Supplementary Figure S4), indicating that pYS can be used together with pAQ-based expression systems. It is known that pCC5.2-derived vector pSCB can coexist with the RSF1010-based vector (Jin et al., 2018), and pSOMA10 derived from pCA2.4 can be harbored with pSOMA16 (derived pCB2.4) or RSF1010-based vectors in *S.* 6803 cells (Opel et al., 2022). The utilization of plasmids in various combinations would allow for complex genetic modifications in cyanobacteria, such as metabolic engineering. Further research on these RCR plasmids will enable the development of more suitable plasmids for material production in cyanobacteria.

Even though pYS4S-GFP does not contain the typical origin of transfer, we have succeeded to obtain a transconjugant of *A.* 7120 that possesses this plasmid and shows GFP fluorescence (Figures 3, 6A). The pYS plasmid has inverted repeat sequences, including the terminator, *lacO* operator, and ColE1, which were assumed to act as the transfer origin of conjugation. The mechanism of conjugation of this plasmid should be clarified in the future. It would be possible to increase the efficiency of conjugation by adding a typical origin of transfer.

Structural prediction indicated that CyRepA is a multi-domain protein (Figure 1B). Interestingly, considering the evolution of the CyRepA protein, CyRepA2 has a DUF3854 domain related to the Toprim domain observed in DNA primase, even though RCR does not require synthesis of a primer RNA. The predicted structure of CyRepA2 is very similar to that of CyRepA1 (Figure 1B). However,

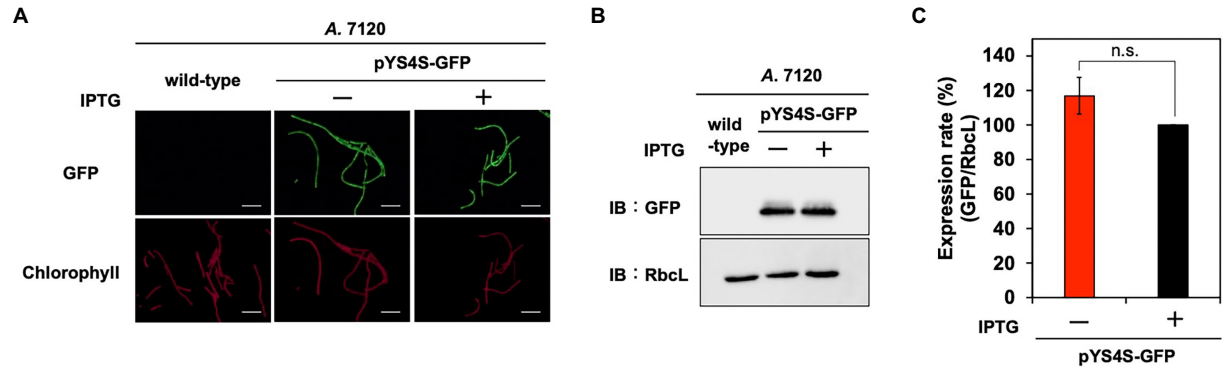

FIGURE 6
Utilization of pYS in *A.* 7120. GFP expression levels in pYS4S-GFP transconjugants were analyzed with (+) and without (–) 1mM IPTG. **(A)** Fluorescence microscopy images. The GFP and chlorophyll images are shown. White bar: 20μm. **(B)** Western blotting analysis. The protein extracts obtained from *A.* 7120 cells were subjected to SDS-PAGE and analyzed by western blotting using antibodies against GFP and control RbcL. **(C)** Comparison of GFP signals. The GFP signals obtained from western blotting analysis were normalized to those of the RbcL signal, and the ratio of the GFP signal in the presence of IPTG was set to 100. Bars represent the mean±SEM (n=3).

the two CyRepAs were compatible with *S.* 6803 cells because pSYSA (coding CyRepA1) and pCC5.2 or pYS1C-GFP (coding CyRepA2) coexist in *S.* 6803 cells (Supplementary Figure S4). In *S.* 7002, the pYS1C-GFP transformant stably retained the pAQ1 plasmid carrying the CyRepA1 homolog, SYNPCC7002_B0001 (Supplementary Figure S4). These observations indicate that CyRepA1 and CyRepA2 function independently as replication factors in cyanobacterial cells. While the size of pSYSA is about 20 times larger than that of pCC5.2, the copy number of pSYSA in *S.* 6803 cells was clearly lower than that of pCC5.2, suggesting that CyRepA1 and CyRepA2 have different properties as replication factors (Nagy et al., 2021). The predicted structures provided important insights into the functional differences between CyRepA1 and CyRepA2. In particular, a clear structural difference between the two was observed in the N-terminal region (Figure 1B). Since CyRepA2 is found in some closely related cyanobacteria, such as *Synechocystis*, *Microcystis*, and *Crocosphaera* (Figure 1A), it may have originated from CyRepA1 in the common ancestor of these species. Future comparisons of CyRepA1 and CyRepA2 will reveal functional differences such as host range and replication activity in cyanobacteria.

Data availability statement

The data presented in the study are deposited in the DDBJ repository, accession number PRJDB11466.

Author contributions

SW: design of this study. YS: data curation, experiment, and writing—original draft preparation. YS, KM, KN-M, TC, and SW: methodology, writing, review, and editing. TC and SW: conceptualization, methodology, formal analysis, supervision, and writing, review, and editing. All authors contributed to the article and approved the submitted version.

Funding

This work was supported by the Ministry of Education, Culture, Sports, Science and Technology of Japan to SW (20K05793) and YS (JP22J13447) and the Advanced Low Carbon Technology Research and Development Program (ALCA) of the Japan Science and Technology Agency (JST; to SW). AR-seq was supported by a Cooperative Research Grant of the Genome Research for BioResource, NODAI Genome Research Center, Tokyo University of Agriculture.

Acknowledgments

We are grateful to Wolfgang Hess for providing valuable information on the function of the Slr7037 protein. We thank Daisuke Umeno and Ryudo Ohbayashi for providing plasmids for constructing pYS1C-GFP and pEX2S-mScarlet, respectively. We sincerely thank Shigeki Ehira for advising us on the transformation of *A.* 7120. We also thank Yuka Kinouchi for her technical assistance.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Supplementary material

The Supplementary material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fmicb.2023.1111979/full#supplementary-material

References

AlQuraishi, M. (2019). AlphaFold at CASP13. *Bioinformatics* 35, 4862–4865. doi: 10.1093/bioinformatics/btz422

Cassier-Chauvat, C., Blanc-Garin, V., and Chauvat, F. (2021). Genetic, genomics, and responses to stresses in cyanobacteria: biotechnological implications. *Gene* 12:500. doi: 10.3390/genes12040500

Castenholz, R. W. (1988). [3] culturing methods for cyanobacteria. *Methods Enzymol.* 167, 68–93. doi: 10.1016/0076-6879(88)67006-6

Chin, T., Okuda, Y., and Ikeuchi, M. (2018). Sorbitol production and optimization of photosynthetic supply in the cyanobacterium Synechocystis PCC 6803. *J. Biotechnol.* 276-277, 25–33. doi: 10.1016/j.jbiotec.2018.04.004

Choi, S. Y., Lee, H. J., Choi, J., Kim, J., Sim, S. J., Um, Y., et al. (2016). Photosynthetic conversion of CO_2 to farnesyl diphosphate-derived phytochemicals (amorpha-4, 11-diene and squalene) by engineered cyanobacteria. *Biotechnol. Biofuels* 9, 1–12. doi: 10.1186/s13068-016-0617-8

Cormack, B. P., Valdivia, R. H., and Falkow, S. (1996). FACS-optimized mutants of the green fluorescent protein (GFP). *Gene* 173, 33–38. doi: 10.1016/0378-1119(95)00685-0

Cramer, P. (2021). AlphaFold2 and the future of structural biology. *Nat. Struct. Mol. Biol.* 28, 704–705. doi: 10.1038/s41594-021-00650-1

del Solar, G., Fernández-López, C., Ruiz-Masó, J., Lorenzo-Díaz, F., and Espinosa, M. (2014). *Rolling Circle Replicating Plasmids. Molecular Life Sciences*. New York: Springer Science

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

Elhai, J., and Wolk, C. P. (1988). Conjugal transfer of DNA to cyanobacteria. *Methods Enzymol.* 167, 747–754. doi: 10.1016/0076-6879(88)67086-8

Englund, E., Andersen-Ranberg, J., Miao, R., Hamberger, B. R., and Lindberg, P. (2015). Metabolic engineering of *Synechocystis* sp. PCC 6803 for production of the plant

diterpenoid manoyl oxide. *ACS Synth. Biol.* 4, 1270–1278. doi: 10.1021/acssynbio.5b00070

Farrokh, P., Sheikhpour, M., Kasaeian, A., Asadi, H., and Bavandi, R. (2019). Cyanobacteria as an eco-friendly resource for biofuel production: a critical review. *Biotechnol. Prog.* 35:e2835. doi: 10.1002/btpr.2835

Felsenstein, J. (1985). Confidence limits on phylogenies: an approach using the bootstrap. *Evolution* 39, 783–791. doi: 10.2307/2408678

Formighieri, C., and Melis, A. (2015). A phycocyanin· phellandrene synthase fusion enhances recombinant protein expression and β-phellandrene (monoterpene) hydrocarbons production in *Synechocystis* (cyanobacteria). *Metab. Eng.* 32, 116–124. doi: 10.1016/j.ymben.2015.09.010

Hasunuma, T., Takaki, A., Matsuda, M., Kato, Y., Vavricka, C. J., and Kondo, A. (2019). Single-stage Astaxanthin production enhances the nonmevalonate pathway and photosynthetic central metabolism in *Synechococcus* sp. PCC 7002. *ACS Synth. Biol.* 8, 2701–2709. doi: 10.1021/acssynbio.9b00280

Higo, A., Isu, A., Fukaya, Y., and Hisabori, T. (2016). Efficient gene induction and endogenous gene repression systems for the filamentous cyanobacterium *anabaena* sp. PCC 7120. *Plant Cell Physiol.* 57, 387–396. doi: 10.1093/pcp/pcv202

Honda, M., Inoue, J., Yoshimasu, M., Ito, Y., Shibata, T., and Mikawa, T. (2006). Identification of the RecR Toprim domain as the binding site for both *recF* and *recO*: a role of *recR* in *recFOR* assembly at double-stranded DNA-single-stranded DNA junctions. *J. Biol. Chem.* 281, 18549–18559. doi: 10.1074/jbc.M512658200

Jin, H., Wang, Y., Idoine, A., and Bhaya, D. (2018). Construction of a shuttle vector using an endogenous plasmid from the cyanobacterium *Synechocystis* sp. PCC6803. *Front. Microbiol.* 9:1662. doi: 10.3389/fmicb.2018.01662

Kaltenbrunner, A., Reimann, V., Hoffmann, U. A., Aoyagi, T., Sakata, M., Nimura-Matsune, K., et al. (2022). Regulation of pSYSA defense plasmid copy number in *Synechocystis* through RNase E and a highly transcribed asRNA. bioRxiv [Preprint]. doi: 10.1101/2022.11.30.518505

Kanesaki, Y., Shiwa, Y., Tajima, N., Suzuki, M., Watanabe, S., Sato, N., et al. (2012). Identification of substrain-specific mutations by massively parallel whole-genome resequencing of *Synechocystis* sp. PCC 6803. *DNA Res.* 19, 67–79. doi: 10.1093/dnares/dsr042

Kato, H., Chibazakura, T., and Yoshikawa, H. (2008). NblR is a novel one-component response regulator in the cyanobacterium *Synechococcus elongatus* PCC 7942. *Biosci. Biotechnol. Biochem.* 72, 1072–1079. doi: 10.1271/bbb.70816

Knoot, C. J., Ungerer, J., Wangikar, P. P., and Pakrasi, H. B. (2018). Cyanobacteria: promising biocatalysts for sustainable chemical production. *J. Biol. Chem.* 293, 5044–5052. doi: 10.1074/jbc.R117.815886

Kuhlemeier, C., Thomas, A., van der Ende, A., van Leen, R. W., Borrias, W., van den Hondel, C. A. M. J. J., et al. (1983). A host-vector system for gene cloning in the cyanobacterium *Anacystis nidulans* R2. *Plasmid* 10, 156–163. doi: 10.1016/0147-619X(83)90068-9

Lan, E. I., and Liao, J. C. (2011). Metabolic engineering of cyanobacteria for 1-butanol production from carbon dioxide. *Metab. Eng.* 13, 353–363. doi: 10.1016/j.ymben.2011.04.004

Letunic, I., Khedkar, S., and Bork, P. (2021). SMART: recent updates, new developments and status in 2020. *Nucleic Acids Res.* 49, D458–D460. doi: 10.1093/nar/gkaa937

Liu, X., Sheng, J., and Curtiss, R. III (2011). Fatty acid production in genetically modified cyanobacteria. *Proc. Natl. Acad. Sci. U.S.A.* 108, 6899–6904. doi: 10.1073/pnas.1103014108

Markley, A. L., Begemann, M. B., Clarke, R. E., Gordon, G. C., and Pfleger, B. F. (2015). Synthetic biology toolbox for controlling gene expression in the cyanobacterium *Synechococcus* sp. strain PCC 7002. *ACS Synth. Biol.* 4, 595–603. doi: 10.1021/sb500260k

Meyer, R. (2009). Replication and conjugative mobilization of broad host-range IncQ plasmids. *Plasmid* 62, 57–70. doi: 10.1016/j.plasmid.2009.05.001

Mirdita, M., Schütze, K., Moriwaki, Y., Heo, L., Ovchinnikov, S., and Steinegger, M. (2022). ColabFold: making protein folding accessible to all. *Nat. Methods* 19, 679–682. doi: 10.1038/s41592-022-01488-1

Mistry, J., Chuguransky, S., Williams, L., Qureshi, M., Salazar, G. A., Sonnhammer, E. L., et al. (2021). Pfam: the protein families database in 2021. *Nucleic Acids Res.* 49, D412–D419. doi: 10.1093/nar/gkaa913

Miyasaka, H., Nakano, H., Akiyama, H., Kanai, S., and Hirano, M. (1998). Production of PHA (poly hydroxyalkanoate) by genetically engineered marine cyanobacterium. *Stud. Surf. Sci. Catal.* 114, 237–242. doi: 10.1016/S0167-2991(98)80750-7

Nagy, C., Thiel, K., Mulaku, E., Mustila, H., Tamagnini, P., Aro, E.-M., et al. (2021). Comparison of alternative integration sites in the chromosome and the native plasmids of the cyanobacterium *Synechocystis* sp. PCC 6803 in respect to expression efficiency and copy number. *Microb. Cell Factories* 20, 1–18. doi: 10.1186/s12934-021-01622-2

Nimura, K., Takahashi, H., and Yoshikawa, H. (2001). Characterization of the *dnaK* multigene family in the cyanobacterium *Synechococcus* sp. strain PCC7942. *J. Bacteriol.* 183, 1320–1328. doi: 10.1128/JB.183.4.1320-1328.2001

Opel, F., Siebert, N. A., Klatt, S., Tüllinghoff, A., Hantke, J. G., and Toepel, J. r., Bühler, B., Nürnberg, D. J., and Klähn, S. (2022). Generation of synthetic shuttle vectors enabling modular genetic engineering of cyanobacteria. *ACS Synth. Biol.* 11, 1758–1771. doi: 10.1021/acssynbio.1c00605

Pettersen, E. F., Goddard, T. D., Huang, C. C., Couch, G. S., Greenblatt, D. M., Meng, E. C., et al. (2004). UCSF chimera—a visualization system for exploratory research and analysis. *J. Comput. Chem.* 25, 1605–1612. doi: 10.1002/jcc.20084

Saitou, N., and Nei, M. (1987). The neighbor-joining method: a new method for reconstructing phylogenetic trees. *Mol. Biol. Evol.* 4, 406–425.

Seki, A., Hanaoka, M., Akimoto, Y., Masuda, S., Iwasaki, H., and Tanaka, K. (2007). Induction of a group 2 sigma factor, RPOD3, by high light and the underlying mechanism in *Synechococcus elongatus* PCC 7942. *J. Biol. Chem.* 282, 36887–36894. doi: 10.1074/jbc.M707582200

Shimada, N., Okuda, Y., Maeda, K., Umeno, D., Takaichi, S., and Ikeuchi, M. (2020). Astaxanthin production in a model cyanobacterium *Synechocystis* sp. PCC 6803. *J. Gen. Appl. Microbiol.* 66, 116–120. doi: 10.2323/jgam.2020.01.003

Tamura, K., Stecher, G., and Kumar, S. (2021). MEGA11: molecular evolutionary genetics analysis version 11. *Mol. Biol. Evol.* 38, 3022–3027. doi: 10.1093/molbev/msab120

Taton, A., Ecker, A., Diaz, B., Moss, N. A., Anderson, B., Reher, R., et al. (2020). Heterologous expression of cryptomaldamide in a cyanobacterial host. *ACS Synth. Biol.* 9, 3364–3376. doi: 10.1021/acssynbio.0c00431

Thiel, T., and Wolk, C. P. (1987). Conjugal transfer of plasmids to cyanobacteria. *Methods Enzymol.* 153, 232–243. doi: 10.1016/0076-6879(87)53056-7

Trieu-Cuot, P., Carlier, C., Martin, P., and Courvalin, P. (1987). Plasmid transfer by conjugation from *Escherichia coli* to gram-positive bacteria. *FEMS Microbiol. Lett.* 48, 289–294. doi: 10.1111/j.1574-6968.1987.tb02558.x

Van der Plas, J., Oosterhoff-Teertstra, R., Borrias, M., and Weisbeek, P. (1992). Identification of replication and stability functions in the complete nucleotide sequence of plasmid pUH24 from the cyanobacterium *Synechococcus* sp. PCC 7942. *Mol. Microbiol.* 6, 653–664. doi: 10.1111/j.1365-2958.1992.tb01513.x

Wang, M., Luan, G., and Lu, X. (2019). Systematic identification of a neutral site on chromosome of *Synechococcus* sp. PCC7002, a promising photosynthetic chassis strain. *J. Biotechnol.* 295, 37–40. doi: 10.1016/j.jbiotec.2019.02.007

Watanabe, S., Ohbayashi, R., Shiwa, Y., Noda, A., Kanesaki, Y., Chibazakura, T., et al. (2012). Light-dependent and asynchronous replication of cyanobacterial multi-copy chromosomes. *Mol. Microbiol.* 83, 856–865. doi: 10.1111/j.1365-2958.2012.07971.x

Wolk, C. P., Vonshak, A., Kehoe, P., and Elhai, J. (1984). Construction of shuttle vectors capable of conjugative transfer from *Escherichia coli* to nitrogen-fixing filamentous cyanobacteria. *Proc. Natl. Acad. Sci. U. S. A.* 81, 1561–1565. doi: 10.1073/pnas.81.5.1561

Xu, W., and McFadden, B. A. (1997). Sequence analysis of plasmid pCC5. 2 from cyanobacterium *Synechocystis* PCC 6803 that replicates by a rolling circle mechanism. *Plasmid* 37, 95–104. doi: 10.1006/plas.1997.1281

Zuckerkandl, E., and Pauling, L. (1965). "Evolutionary divergence and convergence in proteins," in *Evolving Genes and Proteins.* eds. V. Bryson and H. J. Vogel (USA: Elsevier), 97–166.

frontiers | Frontiers in Microbiology

TYPE Original Research
PUBLISHED 21 April 2023
DOI 10.3389/fmicb.2023.1143017

OPEN ACCESS

EDITED BY
Jin Liu,
Peking University, China

REVIEWED BY
Yandu Lu,
Hainan University, China
Stefano Santabarbara,
National Research Council (CNR), Italy

*CORRESPONDENCE
Yong Fan
✉ fanyong@qibebt.ac.cn
Fu-Li Li
✉ lifl@qibebt.ac.cn

RECEIVED 12 January 2023
ACCEPTED 04 April 2023
PUBLISHED 21 April 2023

CITATION
Jiang E-Y, Fan Y, Phung N-V, Xia W-Y, Hu G-R and Li F-L (2023) Overexpression of plastid lipid-associated protein in marine diatom enhances the xanthophyll synthesis and storage.
Front. Microbiol. 14:1143017.
doi: 10.3389/fmicb.2023.1143017

Overexpression of plastid lipid-associated protein in marine diatom enhances the xanthophyll synthesis and storage

Er-Ying Jiang[1,2], Yong Fan[1]*, Nghi-Van Phung[1], Wan-Yue Xia[1], Guang-Rong Hu[1] and Fu-Li Li[1,3,4]*

[1]Shandong Provincial Key Laboratory of Synthetic Biology, Qingdao C1 Refinery Engineering Research Center, Qingdao Institute of Bioenergy and Bioprocess Technology, Chinese Academy of Sciences, Qingdao, China, [2]University of Chinese Academy of Sciences, Beijing, China, [3]Shandong Energy Institute, Qingdao, China, [4]Qingdao New Energy Shandong Laboratory, Qingdao, China

Plastoglobules, which are lipoprotein structures surrounded by a single hydrophobic phospholipid membrane, are subcellular organelles in plant chromoplasts and chloroplasts. They contain neutral lipids, tocopherols, quinones, chlorophyll metabolites, carotenoids and their derivatives. Proteomic studies indicated that plastoglobules are involved in carotenoid metabolism and storage. In this study, one of the plastid lipid-associated proteins (PAP), the major protein in plastoglobules, was selected and overexpressed in *Phaeodactylum tricornutum*. The diameter of the plastoglobules in mutants was decreased by a mean of 19.2% versus the wild-type, while the fucoxanthin level was increased by a mean of 51.2%. All mutants exhibited morphological differences from the wild-type, including a prominent increase in the transverse diameter. Moreover, the unsaturated fatty acid levels were increased in different mutants, including an 18.9–59.3% increase in eicosapentaenoic acid content. Transcriptomic analysis revealed that PAP expression and the morphological changes altered xanthophyll synthesis and storage, which affected the assembly of the fucoxanthin chlorophyll a/c-binding protein and expression of antenna proteins as well as reduced the non-photochemical quenching activity of diatom cells. Therefore, metabolic regulation at the suborganelle level can be achieved by modulating PAP expression. These findings provide a subcellular structural site and target for synthetic biology to modify pigment and lipid metabolism in microalgae chassis cells.

KEYWORDS

plastoglobules, fucoxanthin, eicosapentaenoic acid, fucoxanthin chlorophyll a/c-binding protein, xanthophyll cycles

1. Introduction

Diatoms, as a group of red phytoplankton in the ocean, contribute approximately 20% of the global primary productivity annually and play an essential role in the global carbon fixation and biogeochemical cycle (Field et al., 1998). Diatoms belong to an important branch of photosynthetic organisms. During evolution, prokaryotic cyanobacteria generated two major eukaryotic photosynthetic groups *via* endosymbiosis. One is the green branch, which includes Chlorophyta and land plants that evolved from Prochloron. The other is the red branch, which

evolved from unicellular red algae that underwent endosymbiosis two or three times into various groups such as Cryptophyta, brown algae, diatoms, and dinoflagellates. Four membranes surround the chloroplasts of diatoms because of secondary endosymbiosis, differing from the double-membrane structure of chloroplasts in the common higher plants (Falkowski et al., 2004). The characteristic light-harvesting antenna of diatoms is fucoxanthin chlorophyll a/c-binding protein (FCP), which can capture blue and green light efficiently under low light conditions, thereby maintaining photosystem activity. This is the molecular basis for achieving efficient light energy capture (Wang et al., 2020). Meanwhile, diatoms can quench excessive excitation energy under high light conditions to prevent damage by reactive oxygen species, thereby giving diatoms strong light environment adaptation ability (Büchel, 2014).

As a marine model diatom, *Phaeodactylum tricornutum* has obvious advantages because of its intracellular content of biologically active substances, such as fucoxanthin and eicosapentaenoic acid (EPA). In recent years, fucoxanthin has been confirmed to be a safe and effective dietary supplement with various physiological activities, such as anti-inflammatory, anti-tumor, anti-obesity, and anti-diabetes activities (Maeda et al., 2007; Rokkaku et al., 2013; Martin, 2015). As one of the ω-3 series of polyunsaturated fatty acids (PUFAs), EPA prevents coronary cardiovascular disease and hypertriglyceridaemia and reduces the risk of arteriosclerotic inflammation and various neoplasias (Lorente-Cebrian et al., 2015; Souza and Norling, 2016). The use of biotechnology to improve the levels of fucoxanthin and EPA in *P. tricornutum* has great economic value. In this process, the synthetic pathways of these natural products need to be clarified and optimized.

Plastoglobules (PGs) were originally discovered in the chloroplast lamellae of *Euglena*, and they were characterized as dense osmiophilic globular structures (Wolken and Palade, 1952). Initial studies on PGs revealed that they are formed during the budding of the thylakoid membrane in chloroplasts. PGs attach to the thylakoid membrane and serve as an important site of chloroplast lipid storage, and they possess a single-layered membrane embedded with various proteins (Austin et al., 2006). These structures have spread widely in plastids of different types and developmental times, including chromoplasts and leucoplasts. PGs are relatively large in senescent cells, ranging 0.3–5.0 μm in size (Kaup et al., 2002). Differences in the size and morphology of PGs might serve as a cytological indicator of the growth status and stress tolerance. Among the PGs in the chromoplasts of ripening bell pepper fruits, capsanthin is the most abundant carotenoid, followed by violaxanthin, β-carotene, and capsorubin (Deruere et al., 1994b).

PGs are composed of lipoproteins coated by a monolayer of hydrophobic phospholipid membranes. They have four major components, namely neutral lipids, tocopherols and quinones, carotenoids and their derivatives, and chlorophyll catabolite (van Wijk and Kessler, 2017). According to proteomic analysis, the most abundant proteins in chloroplast PGs are specific members of the plastid lipid-associated protein (PAP)/fibrillin family (pfam04755) and members of the activity of BC1 complex kinase family, which represent approximately 53 and 19% of the PG protein mass, respectively (Singh and McNellis, 2011; Lundquist et al., 2012a). PAP/fibrillin was first isolated from the PGs of sweet pepper in 1994 (Deruere et al., 1994b). Thus far, the PAP/fibrillin family has been recognized as highly conserved, and it is divided into 12 subfamilies in higher plants. The results of subcellular localization experiments revealed that PAP/fibrillin proteins were located in various subcellular organelles of chloroplasts, such as the chloroplast stroma, thylakoid membranes, and PGs (Laizet et al., 2004; Singh and McNellis, 2011; Kim et al., 2018). *Arabidopsis* possesses 14 PAP/fibrillin proteins (termed FBNs in *Arabidopsis* in the following context), 7 of which (FBN1a, FBN1b, FBN2, FBN4, FBN7a, FBN7b, and FBN8) are considered as PG core proteins. The other FBNs are mainly localized in the thylakoid membranes (FBN3a/3b, FBN6, and FBN9) or chloroplast stroma (FBN5) (Lundquist et al., 2012b; Kim E. H. et al., 2015). The functions of FBN1, FBN2, and FBN4, which are localized in PGs, have been validated (Singh et al., 2010; Singh and McNellis, 2011). For example, overexpression of *FBN1a* in tobacco results in an increased number and larger size of PGs, along with increased tolerance to light stress, indicating that this gene functions in responses to biotic stresses (Rey et al., 2000). Knockdown of *FBN1* and *FBN2* in *Arabidopsis* resulted in a similar phenotype as the jasmonate-deficient mutant. Thus, FBN1a, FBN1b, and FBN2 help to recruit jasmonate biosynthetic enzymes to PGs (Youssef et al., 2010). In addition, the PAP/fibrillin family exhibits sequence conservation in the N- and C-terminal regions, including a lipocalin (−like) signature. Based on the presence of a lipocalin (−like) signature in PAP/fibrillin members, they are speculated to contribute to PG function through the binding and exchange of prenyl lipid intermediates. The identified functions of PG core proteins mainly involve the regulation of isoprenoid metabolism and remobilization of thylakoid fatty acids (van Wijk and Kessler, 2017).

Through the proteomic analysis of chromoplast PGs in ripe red peppers, many enzymes related to bicyclic carotenoid biosynthesis have been identified, such as ξ-carotene desaturase (ZDS), lycopene β-cyclase (LCY-β), and β-carotene β-hydroxylase. This suggests that chromoplast PGs have enzymatic functions in carotenoid biosynthesis (Ytterberg et al., 2006). The halotolerant green alga *Dunaliella bardawil* also has PGs with abundant β-carotene content. Analysis of its proteome revealed that it resembles eyespots in *Chlamydomonas reinhardtii* and the chloroplast PGs in *Arabidopsis*. Meanwhile, several enzymes that participate in β-carotene synthesis were identified, including one phytoene synthase gene, two phytoene desaturase genes, two LCY genes, four ZDS genes, and three carotene isomerase genes. Thus, the authors inferred that the abundant β-carotene in *D. bardawil* is probably synthesized in PGs (Davidi et al., 2015).

With continuous research on the structure of PGs, chloroplast PGs are regarded as highly specialized thylakoid microdomains that recruit and concentrate specific proteins and metabolites. In addition, chloroplast PGs play an active role in thylakoid formation, remodeling, and breakdown rather than merely serving a passive storage function as long believed (Brehelin et al., 2007; Besagni and Kessler, 2013). Peter K. Lundquist suggested that PGs are essentially microdomains within the thylakoid membrane, and they likely serve as a platform to recruit proteins and metabolites into spatial proximity, facilitating metabolic channeling or signal transduction to accomplish a series of metabolic functions (Lundquist et al., 2013). PGs comprise a type of microcompartment with integrated roles in plastid metabolism, developmental transitions, and environmental adaptation. Therefore, gene editing of PAP/fibrillin proteins could achieve rational regulation of cell growth metabolism in microalgae.

P. tricornutum is a unicellular organism with unique fucoxanthin synthesis and PUFA accumulation. Previous transmission electron microscopy (TEM) observation revealed that PGs are also present in the chloroplasts of *P. tricornutum*. Research on the pigment and functional protein composition of PGs provides an effective approach for further investigation into the synthesis of different bioactive components in diatoms, hence improving their accumulation.

In this study, a PAP/fibrillin protein in *P. tricornutum* was overexpressed, a large number of mutant strains were obtained, and their phenotypes were validated. We focused on the changes in fucoxanthin content. Concurrently, combined with the photosystem parameters and transcriptomics analysis, emphasis was also placed on the metabolic relationship between PGs and photosystem assembly. We dissected their synthetic regulatory mechanisms and investigated the effects of these mechanisms at the subcellular level.

2. Materials and methods

2.1. Microalgal cultivation

The wild-type (WT) *P. tricornutum* strain was stored at the Microalgae Culture Center (MACC/B228) of Ocean University of China. During autotrophic cultivation, algal cells were cultured in a modified f/2 medium with increasing sodium nitrate concentrations (1 g·L^{-1}). The growth rate of *P. tricornutum* in mixotrophic cultivation can increase significantly. Therefore, 10 g·L^{-1} glycerol was added as the carbon source, and 2 g·L^{-1} tryptone was used as the nitrogen source. All strains were incubated in cell culture flasks and placed in a shaker at 24°C with a rotational speed of 160 r·min^{-1}. The light intensity was 80 μmol photons m^{-2}·s^{-1}.

2.2. Selection of genes and plasmid construction

All *PAP* homologs in *P. tricornutum* and *A. thaliana* were jointly subjected to the phylogenetic tree analysis. Evolutionary analyses were conducted in MEGA11. The evolutionary history was inferred using the maximum likelihood method and Whelan and Goldman + Frequence model. The bootstrap consensus tree inferred from 1,000 replicates represents the evolutionary history of the taxa analyzed. Branches corresponding to partitions reproduced in fewer than 50% of bootstrap replicates were collapsed. The percentage of replicate trees in which the associated taxa clustered together in the bootstrap test 1,000 replicates are presented next to the branches (Tamura et al., 2021). Motifs were analyzed using the online tool MEME (Bailey and Elkan, 1994). Protein domains were predicted using SMART (Letunic et al., 2021). A highly homologous PAP gene (PHATRDRAFT_55153; XP_002184985.1) was finally selected. After sequence cloning, the gene was expressed using the plasmid pPha-T1.

2.3. Electroporation protocol

In total, 2×10^8 cells of *P. tricornutum* during the exponential growth phase were harvested by centrifugation at 1500 × *g* for 10 min at 4°C. After 4–6 washes with 375 mM sterile ice-cold sorbitol, cells were resuspended in 100 μL of 375 mM sorbitol to a final density of 2×10^9 cells·mL^{-1}. Then, a suspension aliquot of 100 μL was mixed with 3–5 μg of DNA linear fragments and 4 μL (10 μg·μL^{-1}) of salmon sperm DNA (denatured by boiling for 1 min), incubated on ice for 10 min, and then transferred into a 2-mm electroporation cuvette. Electroporation was performed using the following settings: 500 V, 400 Ω, and 25 μF (Hu and Pan, 2020). After electroporation, cells were immediately transferred to cell culture flasks containing 10–15 mL of fresh f/2 organic medium and recovered in low light (30 μmol photons m^{-2}·s^{-1}) overnight without shaking. Then, the cells were collected by centrifugation at 1500 × *g* for 10 min and resuspended in 600 μL of fresh f/2 organic medium, and 200 μL of this suspension were plated onto solid medium containing 75 μg·mL^{-1} bleomycin (Zeocin). Then, these plates were placed in an illumination incubator. After 5–7 weeks, the transformants were selected and transferred to liquid f/2 organic medium.

2.4. Determination of the photosynthetic system

The characteristics of the photosynthetic system of microalgae cells were measured as previously described (Ding et al., 2021). Each sample was collected, added to a black 96-well plate, and incubated for 10–15 min in the dark. Two experiments were performed in parallel. Then, the optimal/maximal quantum yield of PSII (Fv/Fm) and non-photochemical quenching (NPQ) were determined using an Imaging-PAM chlorophyll fluorometer (MAXI-Imaging-PAM, WALZ, Germany).

2.5. Rapid detection of fucoxanthin and chlorophyll *a* content in *Phaeodactylum tricornutum*

P. tricornutum cells were collected, and a portion was appropriately diluted. Then, 200 μL of each sample were added to a clear 96-well plate, and the optical density at 750 nm (OD_{750}) was measured by a microplate reader to calculate the turbidity.

Meanwhile, the other portion of *P. tricornutum* cells was harvested by centrifugation at 4,000 × *g* for 5 min, washed once with distilled water, and centrifuged again. The cells were resuspended in ethanol (ethanol:microalgae = 1:1, *v/v*) for 15–20 min. After diluting, 200 μL of each sample were added to a clear 96-well plate, and OD_{445} and OD_{663} were measured using a microplate reader.

The content of fucoxanthin was calculated using the formula (Wang et al., 2018): fucoxanthin content (mg·L^{-1}) = $6.39 \times OD_{445} - 5.18 \times OD_{663} + 0.312 \times OD_{750} - 5.27$.

P. tricornutum cells were collected and centrifuged at 4000 × *g* for 10 min at 4°C. The supernatant was discarded. An equal volume of 90% acetone was added, pipetted, and mixed. The samples were incubated at 4°C in the dark for 60 min. Then, the samples were centrifuged at 4,000 × *g* for 10 min at 4°C, and 200 μL of the supernatant were collected to determine OD_{652} and OD_{655}. The Chl-a content was calculated using the following formula (Lichtenthaler, 1987): Chl-a (mg·L^{-1}) = $16.72 \times OD_{665} - 9.16 \times OD_{652}$.

2.6. Analysis of the pigment composition and fucoxanthin concentration using HPLC

The pigment composition was measured using the Hitachi Primaide HPLC system (Hitachi, Tokyo, Japan) with a C18 reverse phase column (2.7-μm particle size, 100 × 4.6 mm). The mobile phase consisted of acetonitrile and water with a flow rate of 1 mL·min^{-1}. In the gradient condition, the acetonitrile/water ratio was increased from 25:75 to 75:25 over 15 min, maintained for 3 min, and then decreased back to 25:75 over 2 min. The chromatogram was recorded at 445 nm. A fucoxanthin standard (ChromaDex, Irvine, CA, United States) was used to construct a standard curve in the 0.01–1.00 mg•mL^{-1} range.

2.7. Determination of the EPA content by gas chromatography

To analyze the fatty acid profile, *P. tricornutum* cells were harvested by centrifugation at 4000 × *g* for 10 min, washed once with distilled water, and centrifuged again. The samples were placed in a freeze drier and lyophilized overnight, and their dry weight was determined. The lyophilized samples were mixed with chloroform and methanol (*v:v* = 2:1) and then shocked at 45°C for 3 h. Then, KCl (0.9%) was added to the samples, which were centrifuged at 4000 × *g* for 5 min. The bottom layer was transferred to a glass tube and weighed after drying with nitrogen gas. The lipid was added to n-hexane and sulfuric acid methanol (2%), and samples were placed in a baking oven for methyl esterification at 85°C for 2 h. Subsequently, the samples were removed and cooled on ice, and KCl (0.9%) was added. Then, samples were vortexed thoroughly and centrifuged at 4,000 × *g* for 5 min. Finally, the upper organic layer was taken for GC, which was performed using a GC system (7890A, Agilent Technologies, Inc., CA). INNOWAX (30 m × 320 μm × 0.25 μm) was selected as the chromatographic column. In the GC analysis program, an inlet temperature of 250°C and an injection volume of 1 μL were employed. The temperature program settings were 120°C for 5 min, linear ascension at 3.5°C·min^{-1} to 240°C, and a constant temperature for 10 min. N_2 was utilized as the carrier gas with a speed of 28.5 mL·min^{-1} and split ratio of 10:1 (*v/v*).

2.8. Separation of lipid fraction

Polar and non-polar lipids in *P. tricornutum* were separated by using solid phase extraction column (Bond Elut SI, 500 mg/3 mL, Agilent, Santa Clare, CA). The 3 mL silica column was equilibrated by using 3 mL methanol and 9 mL chloroform. Then the lipids extracted from *P. tricornutum* were added into the column. Neutral lipids were eluted with 4.5 mL of chloroform:acetic acid (9:1, *v/v*). Glycolipids were eluted with 6 mL of acetone:methanol (19:1, *v/v*). And phospholipids were eluted with 6 mL of methanol (Damiani et al., 2010). The various lipid fractions were dried with nitrogen gas and added to n-hexane and sulfuric acid methanol (2%). The samples were taken for GC after methyl esterification at 85°C for 2 h as described above Materials and methods 2.7.

2.9. Transcriptomic analysis

On day 14, *P. tricornutum* cells were selected for transcriptomic analysis because the microalgae at this stage had relatively stable morphology and large variations in pigment content. Transcriptome sequencing and analysis were conducted by OE Biotech Co., Ltd. (Shanghai, China). Raw data (raw reads) were processed using Trimmomatic (Bolger et al., 2014). The reads containing ploy-N low-quality reads were removed to obtain clean reads. Then, the clean reads were mapped to the reference genome using hisat2 (Kim D. et al., 2015). The FPKM value of each gene was calculated using cufflinks, and the read counts of each gene were obtained by htseq-count (Trapnell et al., 2010).

2.10. Statistical analysis

All experiments were repeated thrice. Unless otherwise stated, all data are expressed as the mean ± standard deviation. The statistical significance of the values obtained from each experiment was evaluated *via* multiple *t*-tests using GraphPad Prism 8.0.2. Differences were considered significant at $p < 0.05$.

3. Results

3.1. Phylogenetic tree analysis of the PAP genes in *Phaeodactylum tricornutum*

Thirteen homologous fragments of PAP proteins are found in *P. tricornutum*. These sequences were compared with the FBN proteins of *A. thaliana*. A phylogenetic tree was constructed using homology analysis (Figure 1A, Supplementary Data 1). Online analysis of the motifs using the MEME tool revealed that 2–3 fragments of conserved sequences are included in the protein sequence of the PAP/fibrillin family. The structures of all sequences were predicted and visualized using the SMART batch function in Tbtool software (Chen et al., 2020). We found that almost all of the PAP proteins in *P. tricornutum* have a signal peptide sequence or transmembrane domain. Among them, PHATRDRAFT_48066 and PHATRDRAFT_41069 have significant transmembrane domains. Excluding PHATRDRAFT_33731, all PAP proteins contained a chloroplast signal peptide. Meanwhile, PAP proteins in *A. thaliana* did not possess signal peptides and transmembrane domains. Previously, no PG protein was known or predicted to possess transmembrane domains, consistent with PGs being bound by a membrane lipid monolayer (van Wijk and Kessler, 2017). The presence of a signal peptide domain in *P. tricornutum* PGs might be related to their four-layered chloroplast membrane structure. Only a few PAP proteins of *P. tricornutum* are clustered into the same branch as the FBNs of *A. thaliana*. Among them, PHATRDRAFT_55153 and PHATRDRAFT_45813 clustered into one branch with AT4G00030 and AT3G58010, respectively, with bootstrap consensus values exceeding 80. Further combined with conserved sequence analysis, PHATRDRAFT_55153 has a more consistently conserved sequence similarity. Therefore, PHATRDRAFT_55153 was overexpressed in *P. tricornutum* to analyze the changes in PGs and their function in cell metabolism. The protein structure was predicted by Alphafold 2

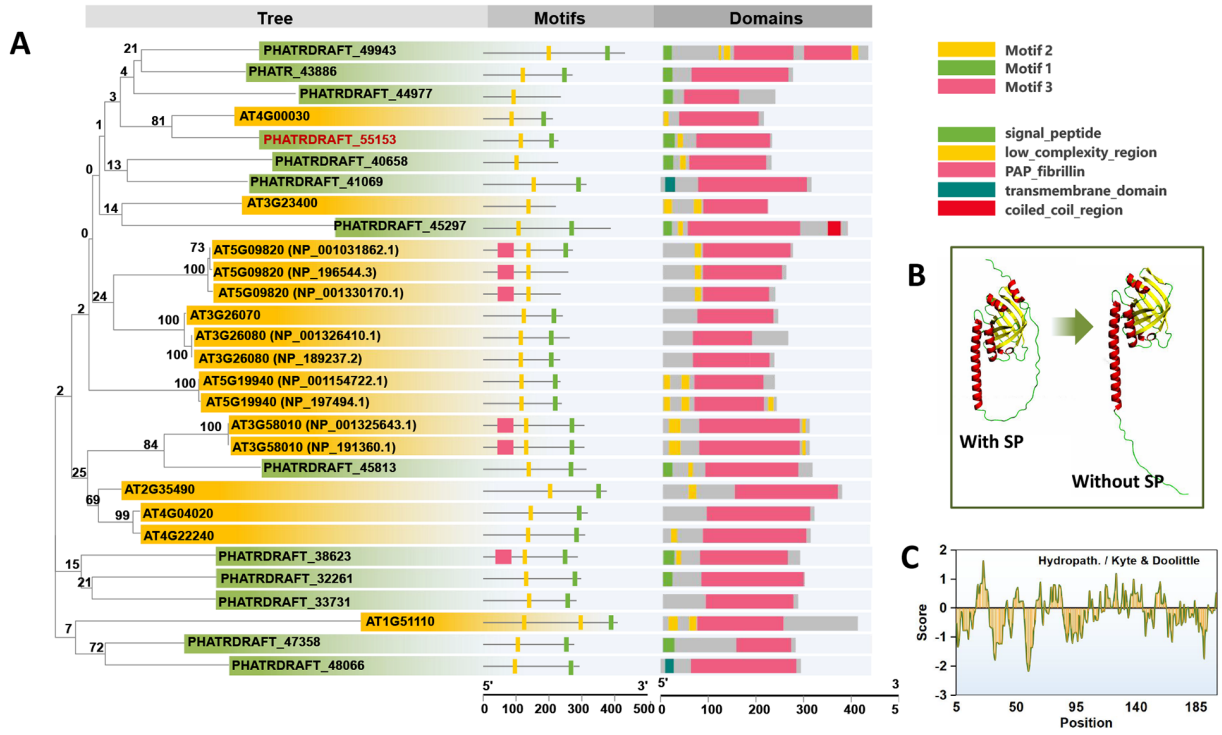

FIGURE 1
Sequence analysis of PAP proteins from *P. tricornutum*. **(A)** Phylogenetic tree analysis, motif analysis, and functional domain analysis of PAP/fibrillin proteins in *P. tricornutum* and *A. thaliana*. **(B,C)** Protein structure prediction and hydrophobicity analysis of PHATRDRAFT_55153. SP: signal peptide. All protein sequences are provided in Supplementary Data 1.

(Jumper et al., 2021). The results illustrated that after the signal peptide removal, the protein had a long and straight N-terminus, and the C-terminus formed an eight-stranded antiparallel beta-barrel. Such structural and hydrophobic region characteristics might play special roles in the formation and maintenance of the monolayer structure (Figures 1B,C).

3.2. Overexpression of the PAP gene leads to morphological changes

Endogenous *PAP* was overexpressed by electroporation. Gene insertions were achieved by random integration. Among the obtained mutant strains, six strains (named PAP-A ~ F) were selected for subsequent analysis. At the initial stage of cultivation, microscopic observation revealed apparent differences in cell morphology, particularly an increase in transverse diameter and a decrease in the longitudinal diameter between the mutant and WT strains. The mutant strain PAP-B displayed a triangular cell morphology. Statistical analysis was performed in 15 random microscopic fields (six mutants were selected for observation, and the cell density was 6×10^7 cells•mL^{-1} during observation). The mutants were predominantly larger than the WT in the transverse diameter. In particular, the transverse diameter was more than 10% larger than that of the WT in 92.3% of the mutants. Triangular-shaped cells accounted for 85.5% of the total cells of PAP-B at the initial stage of cultivation (Figures 2A–C). We speculate that overexpression of *PAP* affects the cell wall and cytoskeleton, thereby causing changes in cell morphology. This phenomenon was neither expected nor has it been reported.

Furthermore, PAP-B and PAP-C, which had the most extensive morphological differences among the mutant strains, were selected for TEM observation of PG structures (Figures 2D,E). First, the numbers of PGs did not significantly differ among different strains because of the differences in the choice of cell section. Different from the number of PG structures, the size of PGs was not affected by the choice of cell section. Therefore, we compared the diameter between the mutants WT, observing that the diameters of PGs were smaller in the mutants (Figures 2F,G). This phenomenon supports the existence of a phenotypic association between the predicted functional protein PHATRDRAFT_55153 and the PGs. The expression of this PAP protein might affect the size and structure of PGs. Further validation of the protein function might require immunoelectron microscopy-based approaches for quantitative analysis.

3.3. An essential role of the PAP gene in photosystem assembly

The mutants and WT were incubated with shaking under the same mixotrophic conditions. Comparing the growth curve by measuring OD_{750}, the mutant strains grew slower than the WT. With increasing cell passage, the differences in growth rates almost disappeared. The cell growth curves after three passages (generation time = 20 days) are presented in Figure 3A. The change in growth rates was synchronous with the aforementioned differences in cell

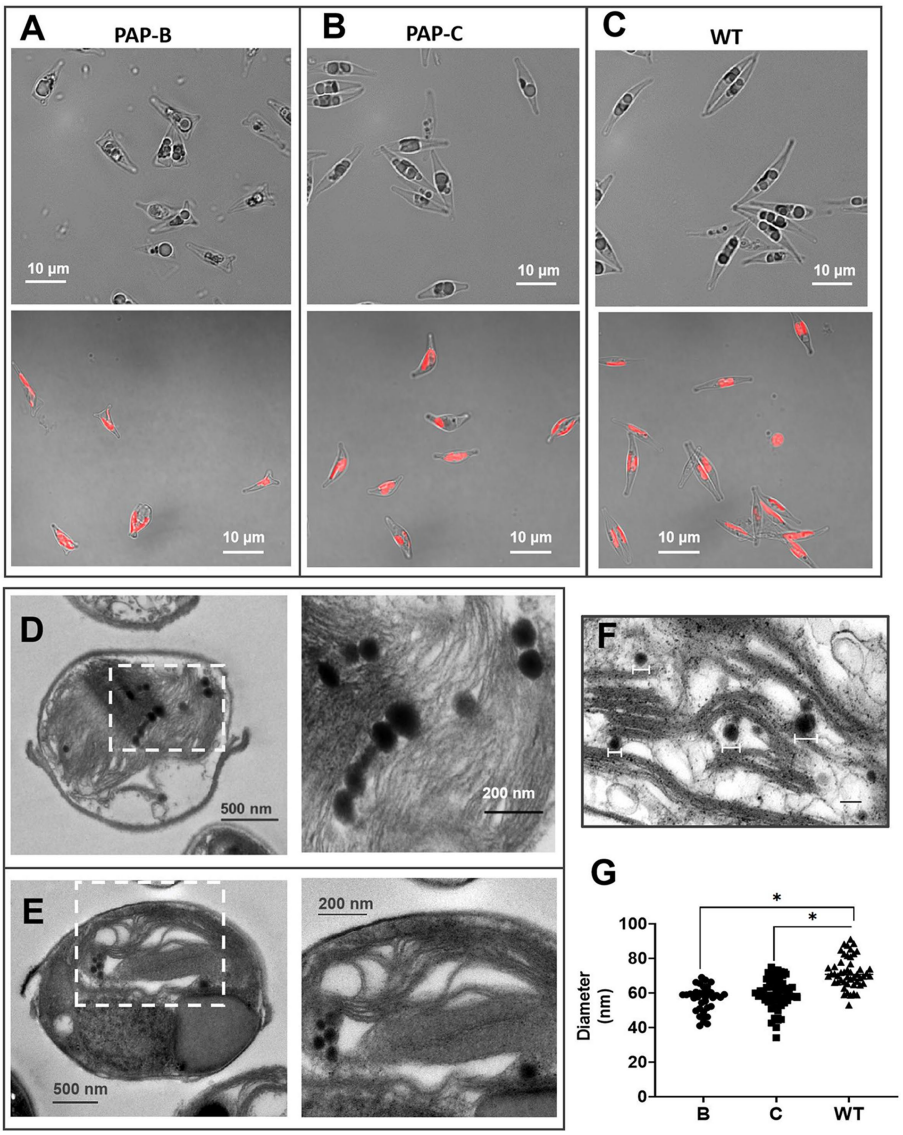

FIGURE 2
Morphological analysis of the mutants and WT. **(A–C)** Micrographs of PAP-B, PAP-C, and WT. The bottom images present the distribution of spontaneous chlorophyll fluorescence in cells. **(D,E)** Transmission electron microscopy (TEM) images of the WT and mutants. The right image is a partial enlargement of the left image. **(F)** ImageJ was used to calculate the diameters of PGs in the TEM images. **(G)** Statistical analysis of the PG diameters in different cells.

morphology. PCR-based analysis of the genome found no loss of the transformed gene fragments. Changes in the cell morphology and growth rate might have resulted from adaptive evolution.

The photosynthetic parameters of the mutants and WT were determined during growth. There was no significant difference in Fv/Fm between the mutants and WT (Figure 3B). However, in the photosynthetic kinetics analysis, NPQ was significantly larger in the WT than in the mutants, whereas Y(NO) was lower in the WT than in the mutants (Figures 3C,D). These results indicate that the WT can more actively shield the photosystem from destruction under high light conditions. The mutants also had a slightly higher photosynthetic electron transport rate (ETR) and Y(II) than the WT (Figures 3E,F). This series of changes in photosynthetic parameters illustrate that the efficiency of the photochemical reaction center was not affected, whereas the changes mainly focused on light-harvesting antenna proteins, especially heat dissipation-associated molecules. NPQ is one of the most rapid mechanisms diatoms possess to dissipate excess energy. Its capacity is mainly defined by the xanthophyll cycle (XC) and light-harvesting complex X (Lhcx) proteins (Blommaert et al., 2020). The diadinoxanthin (Ddx) de-epoxidation is one of the fastest biochemical responses of the thylakoid membrane to environmental factors (Bojko et al., 2019). Previous studies described a carotenoid biosynthesis pathway in PGs, implying that PGs in chloroplasts are the essential structures for carotenoid biosynthesis (Deruere et al., 1994a; Vishnevetsky et al., 1999; Ytterberg et al., 2006; Singh and McNellis, 2011). We speculate that the changes in PGs associated with PAP genes affect the synthesis of intracellular xanthophylls that bind photosynthetic antenna proteins, which will significantly affect the efficiency of NPQ. Therefore, we further analyzed the contents of photosynthetic pigments.

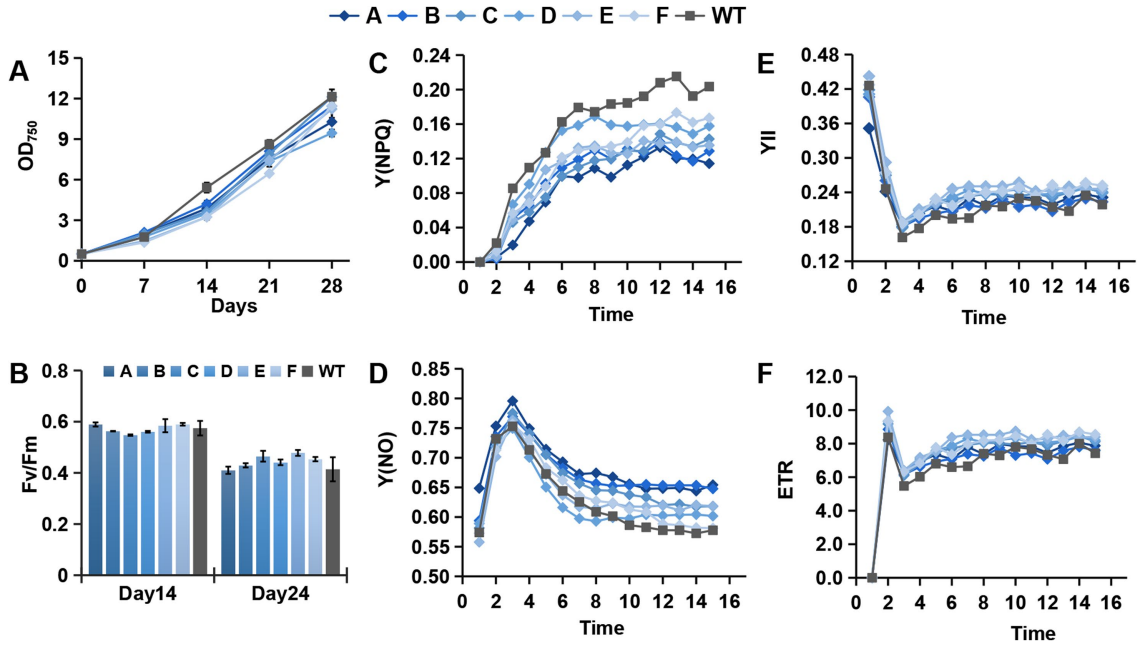

FIGURE 3
Growth and photosynthetic parameters of different mutants and the WT. **(A–F)** Growth curve, Fv/Fm, Y(NPQ), Y(NO), YII, and ETR of the photosystem.

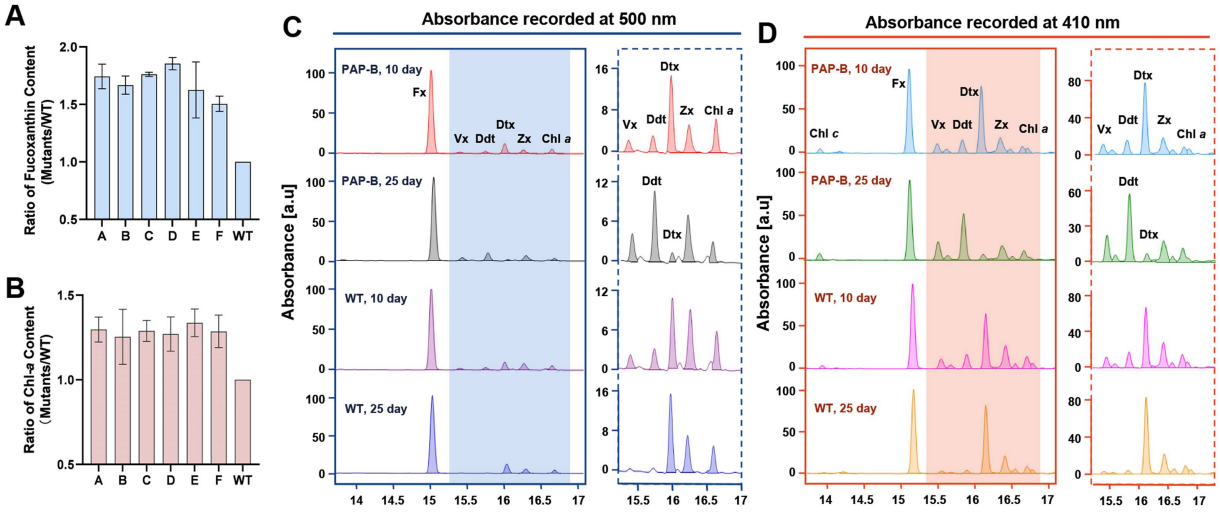

FIGURE 4
Pigment composition of the mutant PAP-B and WT. **(A,B)** The ratio of fucoxanthin and Chl-a content between the mutant and WT after 14days. **(C,D)** The pigment compositions of PAP-B and WT were analyzed by HPLC at different culture times, and the absorption peaks at 500 and 410nm are presented.

3.4. Overexpression of the PAP gene improved fucoxanthin and Chl-*a* levels

In this study, the intracellular content of fucoxanthin during cultivation was analyzed. The results illustrated that the fucoxanthin content was higher in the mutants than in the WT in different cultivation periods. The PAP-D strain exhibited 85.39% higher fucoxanthin content than the WT after 14 days of cultivation (Figure 4A). On day 24 of cultivation, the fucoxanthin content of mutants was on average 33.10% higher than that of the WT. In addition, the intracellular content of Chl-*a* was analyzed. The trend of Chl-*a* content was consistent with that of fucoxanthin. The mutants had an average 28.85% higher Chl-*a* level than the WT after 14 days of culture (Figure 4B). On day 24, this difference decreased to an average of 15.62%. The increase in Chl-*a* content also made the PAP-B cells appear dark green in a short period rather than reflecting the brown color of diatoms. The expression of the predicted PAP protein resulted in enhanced fucoxanthin and chlorophyll levels in the

mutants. In *P. tricornutum*, the crystal structure of FCP reveals that the Lhcf4 protein binds seven Chl-*a*, two Chl-*c*, seven fucoxanthin, and probably one Ddx moiety (Wang H. et al., 2019). The assembly of FCPs and pigments into a complex, followed by further assembly with a photochemical reaction center into a super complex, is responsible for converting light energy into chemical energy (Pi et al., 2019; Wang et al., 2020). However, the proportion of this pigment content is not constant, as the xanthophyll content in particular can vary with the growth environment. Significantly, marine algae experience non-periodic fluctuations in their exposure to light because of water mobility. In diatoms, NPQ is associated with the transformation and accumulation of the XC pigments Ddx and diatoxanthin (Dtx) (Büchel, 2020; Kuczynska et al., 2020). Their concentrations determine the magnitude of the NPQ response. Under high light, FCPs increase their XC pigment content and the de-epoxidation ratio (Lepetit et al., 2012), thereby enhancing their capacity to dissipate energy (Gundermann and Büchel, 2012). Enrichment of Dtx in PSII under high light suggests that photosystems are organized to allow xanthophylls to remain in dynamic balance. This means that xanthophylls are not synthesized and decomposed continuously, but require a storage structure near the photosystem. Combined with the change of PG structures, we speculate that *PAP* overexpression leads to the formation of a more high-density isopentene microcompartment, which affects the exchange of xanthophylls on the thylakoid membrane and PGs, thus changing the assembly of the photosystem.

When the pigment composition of the mutant PAP-B and the WT was analyzed by HPLC, a significant difference was observed in the levels of Ddx and Dtx. The Ddx level was significantly higher than that of Dtx in PAP-B at the later stage of culture. In contrast, the Ddx level was much lower than the Dtx level in PAP-B during the early stage and in the WT during different periods. Ddx should be de-epoxidated to Dtx under light on the thylakoid membrane (Figures 4C,D), whereas the mutant strain at the later stage cultured under the same conditions had a high level of Ddx, most likely because Ddx cannot be integrated into thylakoid membranes nearby the light-harvesting antenna proteins and therefore it was not de-epoxidated. Combined with the previous speculation, a significant amount of Ddx might be "trapped" in the PGs without transport into the photosynthetic system complex.

3.5. Transcriptome data analysis

The transcriptome data of PAP-B, which featured the most significant change in cell morphology, were compared with those of the WT. We selected the cells cultured on day 14 for transcriptomic analysis because the morphology of microalgae cultured to this stage was relatively stable and the difference in the pigment content reached its maximum.

First, we focused on the expression of all annotation genes in the carotenoid synthesis pathway. The most strongly upregulated gene was *CRTISO4* ($\log_2$FC = 2.57), which encodes carotenoid isomerase. Recently, this gene was also proven to participate in the cis-trans isomerization of phytoene, a key gene for lycopene formation (Sun et al., 2022). In addition, most of the enzymes in the fucoxanthin synthesis pathway were upregulated (Figure 5A), which was consistent with increased fucoxanthin content. In addition, XC-related enzymes were also significantly upregulated. The transcriptome data also revealed significant upregulation of the violaxanthin de-epoxidase-like gene (PHATRDRAFT_46155). The upregulation of this enzyme is usually accompanied by stress, which is required to initiate the Ddx cycle on the thylakoid membrane (Goss and Latowski, 2020).

The most strongly downregulated gene was carotenoid cleavage dioxygenase, which can catalyze the oxygenolytic fission of alkene bonds in carotenoids to generate apocarotenoid products (Zhou et al., 2019). The downregulation of degradation pathways in the mutant strains was accompanied by the accumulation of carotenoids.

Transcriptomic analysis of the chlorophyll synthesis and degradation pathways in mutants revealed that most of the key genes of the chlorophyll synthesis pathway were upregulated. In contrast, most of the genes of the chlorophyll degradation pathway were downregulated. In particular, the pheophytin pheophorbide hydrolase (pheophytinase, *PPH*) gene was significantly downregulated. PPH is a critical enzyme in Chl degradation. Mutagenesis or overexpression of *PPH* can lead to a stay-green or premature senescence phenotype in *Arabidopsis* and rice (Schelbert et al., 2009; Wang W. et al., 2019). Previous studies identified PPH in isolated PGs, indicating that PPH is likely a *bona fide* PG protein (Lundquist et al., 2012b).

Further analysis of the transcriptome data demonstrated that most antenna proteins were downregulated (Figure 5B). Previous studies on the FCP crystal structure in *P. tricornutum* revealed that functional PSII–FCPII monomers include one PSII core, two FCP tetramers, and three FCP monomers (FCP-D/E/F). One of the FCP tetramers is directly associated with the core at the CP47 side and designated strongly associated tetramer [S-tetramer (ST)], whereas the other one is associated with the PSII core indirectly at the CP43 side through two FCP monomers, FCP-D and FCP-E, and hence designated moderately associated tetramer [M-tetramer (MT)] (Wang et al., 2020). The transcriptome results for FCP complexes illustrated that only a few antenna proteins were upregulated, and most FCP antenna proteins were downregulated. This might lead to the inhibition of antenna protein assembly.

PG structures enable the synthesis and storage of intracellular carotenoid substances. The protein structural properties of PAP might increase the ability of PGs to store and bind carotenoid substances, thus reducing carotenoid transport into the photosynthetic system. This is responsible for the reduced synthesis of the corresponding antenna proteins during the assembly of the photosynthetic system.

Lundquist speculated that PGs are essentially microdomains within the thylakoid membrane that potentially serve as platforms to recruit proteins and metabolites to facilitate metabolic channel activity or signal transduction (Lundquist et al., 2013). We further suggest that PGs act as a pool, and alterations in PAP protein expression might affect pigment metabolism in the thylakoid membrane, resulting in unexpected modifications in photosynthetic regulation.

Due to the NPQ difference between mutants and WT, we also detected the expression of *Lhcx* genes in the photosynthetic system. Diatoms possess an impressive capacity of NPQ, provided by the xanthophyll diatoxanthin and Lhcx proteins, and there are four *Lhcx* genes in *P. tricornutum* (Buck et al., 2019). The results showed that *Lhcx2* in the mutant strain was significantly downregulated, while the expression of *Lhcx3* was upregulated considerably (Figure 5B). However, the expression level of *Lhcx3* in different strains is very low (Supplementary Data 2). In this case, the difference in NPQ between the mutant and WT may be more influenced by *Lhcx2* and the Ddx/Dtx ratio.

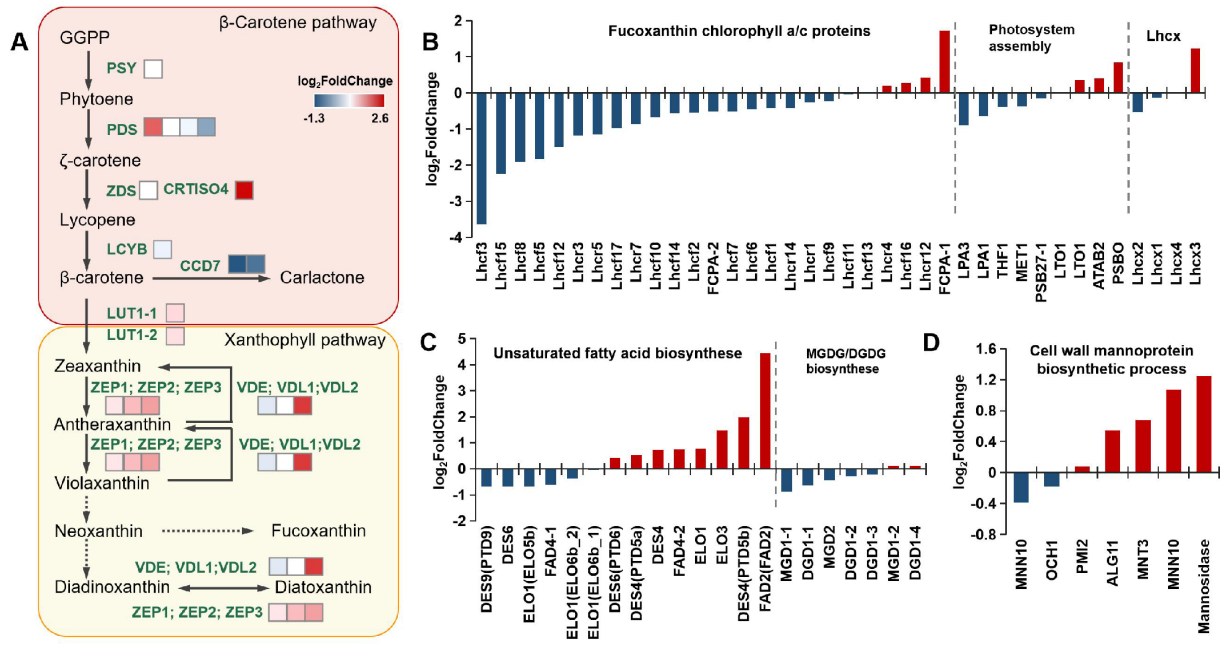

FIGURE 5
Transcriptomic analysis of different metabolic pathways. **(A)** Expression of genes in the carotenoid synthesis pathway. **(B)** Expression of genes in the FCP complex, photosystem assembly, and the *Lhcx* genes. **(C)** Expression of genes in the polyunsaturated fatty acid synthesis pathway and MGDG/DGDG synthesis pathway. **(D)** The mannose metabolism genes are related to cell wall synthesis. All abbreviations and data are provided in Supplementary Data 2.

TABLE 1 Fatty acid composition of the WT and transformants.

Strains	EPA (%)	SFA[a] (%)	MUFA[b] (%)	PUFA[c] (%)
A	19.82 ± 0.02	31.37 ± 0.22	30.52 ± 0.49	38.1 ± 0.71
B	22.12 ± 0.15	31.86 ± 0.13	29.06 ± 0.46	39.07 ± 0.33
C	21.48 ± 0.13	38.26 ± 0.07	26.12 ± 0.45	35.60 ± 0.52
D	20.21 ± 0.25	35.63 ± 0.39	28.13 ± 0.21	36.23 ± 0.59
E	24.5 ± 0.13	30.64 ± 0.07	24.15 ± 0.05	45.20 ± 0.02
F	18.29 ± 0.13	34.40 ± 0.33	31.96 ± 0.18	33.63 ± 0.49
WT	15.38 ± 0.54	41.19 ± 1.01	31.27 ± 0.12	27.52 ± 1.13

[a]indicate the saturated fatty acids, monounsaturated fatty acids, and polyunsaturated fatty acids, respectively.
[b]indicate the saturated fatty acids, monounsaturated fatty acids, and polyunsaturated fatty acids, respectively.
[c]indicate the saturated fatty acids, monounsaturated fatty acids, and polyunsaturated fatty acids, respectively.

In addition to carotenoid and photosystem metabolism, we performed transcriptomic analysis of lipid metabolism in the mutants (Figure 5C). The chloroplast thylakoid membrane is mainly composed of lipids and different protein–pigment complexes. The hydrophobic region inside the membrane bilayer interacts with membrane proteins to ensure that the light energy absorbed by the pigment causes the production and transmission of photosynthetic electrons and the proper progression of photophosphorylation. The major lipid type in the thylakoid membranes of photosynthetic organisms is polar glyceride, which includes three glycolipids [monogalactosyl diacylglycerol (MGDG), digalactosyl diacylglycerol (DGDG), and sulfoquinovosyl diacylglycerol (SQDG)] and phospholipid (phosphatidyl glycerol). Among these glycerides, MGDG and DGDG, which contain high proportions of PUFAs, accounted for approximately 50 and 30% of the total lipid content, respectively (Goss and Latowski, 2020). A high proportion of PUFAs is believed to be related to the high fluidity of the membrane, and it is an essential feature for efficient photochemical reactions on thylakoid membranes. In addition, 20 DGDG, 42 MGDG, 16 SQDG, and 30 phosphatidylglycerol molecules are found in a PSII–FCPII dimer. These lipids are mostly distributed in the interfaces between subunits, suggesting their roles in mediating subunit interactions (Wang et al., 2020). In the transcriptomic analysis, most genes of the PUFA synthesis pathway were upregulated, especially delta (12)-fatty-acid desaturase 2. However, the expression of enzymes synthesizing MGDGs and DGDGs did not significantly differ (Figure 5C). We analyzed the fatty acid composition of the mutants and WT using GC. The results indicated that the proportion of EPA present in each mutant was significantly higher than that in the WT. This was directly linked to the significant upregulation of fatty acid desaturases identified from the transcriptomic analysis (see Table 1).

By isolating lipids based on their polarity, we purified the neutral lipids, glycolipids, and phospholipids for GC analysis. The results revealed that PUFAs such as EPA levels of the mutants were significantly higher than that of the WT in the fatty acid composition of glycolipids. In contrast, no significant differences were found in neutral lipids and phospholipids (Supplementary Figure 1; Supplementary Table 1). PUFAs such as EPA mainly accumulate in membrane lipids. This indicates the increase of membrane fluidity during the assembly of the photosynthetic system. Consequently, the interactions between the thylakoid membranes and the PS–FCP subunits would be induced.

Finally, we focused on cell wall-related metabolism, as significant differences in cell morphology were observed upon phenotype comparisons. The cell wall of *P. tricornutum* has low silica content, differing from other diatoms, which are mainly composed of organic molecules, notably sulfated glucuronomannan (Le Costaouëc et al., 2017). The polysaccharide backbone consists of a mannan chain decorated with sulfate ester. It was proposed that the branching consists of mannose and glucuronic acid based on structural analyses of fragments obtained by mild acid hydrolysis. We mainly analyzed mannose metabolism genes involved in the cell wall synthesis pathway. The results indicated that the related genes (alpha-mannosidase, alpha-1,6-mannosyltransferase, alpha-1,2-mannosyltransferase) were upregulated (Figure 5D). Because the cell wall synthesis pathway in *P. tricornutum* is unclear, there was no differential comparison of other annotated genes related to cell wall synthesis.

Previous studies on different morphological cells of *P. tricornutum* illustrated that the cell wall of *P. tricornutum* exhibits high plasticity compared to those of other diatoms, and it can display three morphotypes: fusiform, oval, and triradiate. In comparison, the cellular morphology of our mutant strains appeared to be intermediate between fusiform and oval. *P. tricornutum* cells will undergo different physiological changes upon exposure to environmental stress. In particular, comparative transcriptomic analysis based on EST indicated that the oval morphotype features the upregulation of genes encoding proteins involved in hyposalinity and/or cold stress responses. Meanwhile, the plasticity of the cell wall is also sensitive to different stress conditions (Vartanian et al., 2009; Ovide et al., 2018; Galas et al., 2021).

4. Discussion

Chloroplast PGs are dynamic monolayer membrane structures containing special metabolites and proteins. They store secondary metabolites such as pigments and play an active role in the developmental transition and environmental adaptation, making them microcompartments with integrated functions. A comprehensive functional model of PGs was constructed by co-expression analysis using PG proteins, and the co-expression network was involved in four specific functions of PGs: (1) senescence, (2) plastid biogenesis, (3) isoprenoid lipid metabolism, and (4) redox/photosynthetic regulation (Lundquist et al., 2012b).

This study obtained a series of mutants by overexpressing a predicted PAP gene. We unexpectedly found that the cell morphology of the mutants all shifted toward an oval morphotype. In phenotype detection, the unified changes of the mutants typically included increases in the levels of fucoxanthin and PUFAs, especially EPA, and a decrease in NPQ activity. PGs are located on the thylakoid membrane and extended out by the thylakoid membrane. Therefore, the changes in PGs and cells were caused by *PAP* overexpression, which we believe is closely related to the assembly and function of the photosystem (Figure 6).

First, the diameter of PGs became smaller after *PAP* overexpression. Combined with the previous model, changes in PG inclusions affect PG size (Lundquist et al., 2013). The nonpolar components are buried inside the PGs and covered by polar lipids and proteins on the surface of PGs. The protein structure of PAP renders it amphipathic. Its N-terminal α-helix can bind to the polar lipid of the monolayer, and the β-sheets of the tail constitute a barrel structure that could bind relatively hydrophobic substances. The ratio of amphiphilic-to-hydrophobic metabolites in PGs is reflected in their surface area/volume ratio and thus the diameter of the approximately spherical PGs. When the overexpressed PAP protein is located on the PG surface of *P. tricornutum*, more surface sites are occupied by the protein, which might more specifically bind to isoprenoids inside the PGs. Consequently, isoprenoids are more strongly accumulated in the interior of PGs than other metabolites, and this selectivity might lead to a decrease in the spherical volume. In fact, we suggest that the structural properties of this PAP protein determine the preference of PGs for the encapsulated contents.

In diatoms, a Ddx–Dtx cycle exists in the FCP antennae, and it quenches excess energy under strong light conditions. In our study, the photosynthetic electron transport efficiency and photosynthetic efficiency of the mutants did not show an obvious decrease, whereas NPQ was decreased. We speculated that increases in PAP protein levels on PGs and their binding capacity for isopentenyls affected the storage and transport efficiency of xanthophylls. Consequently, the exchange efficiency of substances between PGs and photosynthetic systems and the dynamic equilibrium of xanthophylls in these systems were altered. Regulating the number of PGs and the expression of PAP proteins appears to play a role in pigment metabolism in the photosynthetic system. We suggest that PGs act as buffering pools for carotenoids in plant cells, making them essential for regulating the photosynthetic efficiency of diatoms.

To use *P. tricornutum* as an industrial microorganism for production, unlike the natural environment in which microalgae need to adapt to constant changes in light and temperature; in the production process, especially in mixotrophic fermentation, the environment of the algal cells is relatively stable regarding light and temperature. Under this circumstance, engineering algal species to obtain strains that efficiently accumulate fucoxanthin at the expense of reduced redundant NPQ capacity can be used as a strategy for algal species modification in the future.

EPA is another high-value product of *P. tricornutum*. The content of this PUFA was also elevated in the mutant strains. We also believe that there is a close relationship between the accumulation of EPA and the assembly of the photosynthetic system of the cell. In the membrane lipids of thylakoids of diatoms, MGDG acts as a solvent for XC pigments (Latowski et al., 2004). EPA preferentially binds to the sn-1 position of the glycerol backbone in MGDGs. DGDGs exhibit a fatty acid component comparable to that of MGDGs (Dodson et al., 2014). Similar to MGDG, EPA generally occupies the sn-1 position of DGDGs. In thylakoid membranes, MGDG appears to play a key role in providing membrane fluidity, which is essential for the efficient diffusion of XC pigments (Goss and Latowski, 2020).

Finally, in terms of changes in cell morphology, we suggest that both the synthesis and metabolism of the cell wall affect cell morphology. These processes are also associated with alterations in the cellular content and cytoskeleton. Many studies have examined the cell wall formation of diatoms and its alteration with changes in the spectra (Vartanian et al., 2009). The morphology of *P. tricornutum* is easily comparable to that of a convex lens. The cell morphology and its contents cause light refraction, which is consistent with the distribution of the photoreactive centers of chloroplasts and the advantages of the FCP complex in absorbing spectra of different wavelengths. In our study, the morphological

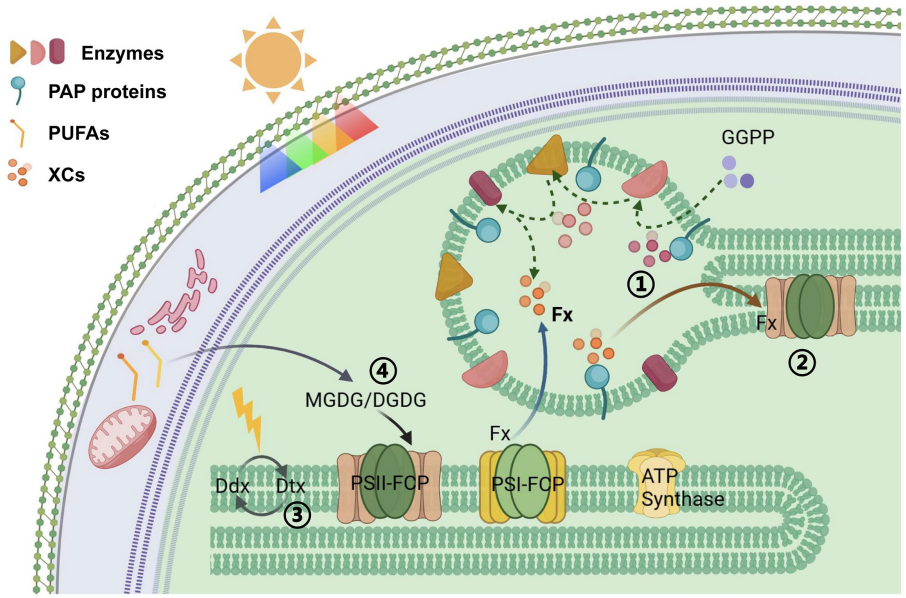

FIGURE 6
Schematic overview of proposed organization: Overexpression of the *PAP* gene (1) led to the accumulation of xanthophylls in PGs, (2) affected the assembly of photosynthetic system antenna proteins, (3) reduced the efficiency of XC in thylakoids, (4) increased the degree of unsaturation of MGDG and DGDG in thylakoid membranes, thereby enhancing membrane fluidity.

changes of the mutants might increase the refractive index of the "convex lens," thus having a great refractive effect on the spectrum. As an explanation, we tend to attribute this phenomenon to the lack of lutein-like substances dependent on the assembly of the FCP complex, resulting in the dependence on specific spectra. In addition, we found that transitions between multiple morphologies of *P. tricornutum* are probably attributable to the spectral dependence of the photosystem.

PGs are structures of subcellular organelles. The single-layered membranous structure formed on the thylakoid gives it close contact with the photosynthetic system for the exchange of substances. Our study preliminarily revealed the role of PGs as a pool for synthesizing and storing xanthophylls in photosynthetic systems in a primitive marine diatom. This finding inspires the utilization of PGs in metabolic engineering. This subcellular organelle structure can be modified and engineered as a functional block, thus providing a probable target site for synthetic biology.

Data availability statement

The datasets presented in this study can be found in online repositories. The names of the repository/repositories and accession number(s) can be found in the article/Supplementary material.

Author contributions

E-YJ, YF, and W-YX accomplished the gene isolation, transformation, positive transformants screening, and carotenoid analysis. G-RH and F-LL supervised the experiments and provided guidance. W-YX and N-VP contributed to the fatty acid analysis. YF and E-YJ designed the experiment, analyzed the data, and drafted the manuscript. All authors contributed to the article and approved the submitted version.

Funding

This work was supported by the National Key Research and Development Program of China (2021YFA0909703; 2019YFD0901904), The National Natural Science Foundation of China (31973007), Shandong Energy Institute (SEI I202136).

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

Supplementary material

The Supplementary material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fmicb.2023.1143017/full#supplementary-material

References

Austin, J. R., Frost, E., Vidi, P. A., Kessler, F., and Staehelin, L. A. (2006). Plastoglobules are lipoprotein subcompartments of the chloroplast that are permanently coupled to thylakoid membranes and contain biosynthetic enzymes. *Plant Cell* 18, 1693–1703. doi: 10.1105/tpc.105.039859

Bailey, T. L., and Elkan, C. (1994). Fitting a mixture model by expectation maximization to discover motifs in biopolymers. *Proc. Int. Conf. Intell. Syst. Mol. Biol.* 2, 28–36.

Besagni, C., and Kessler, F. (2013). A mechanism implicating plastoglobules in thylakoid disassembly during senescence and nitrogen starvation. *Planta* 237, 463–470. doi: 10.1007/s00425-012-1813-9

Blommaert, L., Vancaester, E., Huysman, M. J. J., Osuna-Cruz, C. M., D'hondt, S., Lavaud, J., et al. (2020). Light regulation of LHCX genes in the benthic diatom Seminavis robusta. *Front. Mar. Sci.* 7:e00192. doi: 10.3389/fmars.2020.00192

Bojko, M., Olchawa-Pajor, M., Goss, R., Schaller-Laudel, S., Strzalka, K., and Latowski, D. (2019). Diadinoxanthin de-epoxidation as important factor in the short-term stabilization of diatom photosynthetic membranes exposed to different temperatures. *Plant Cell Environ.* 42, 1270–1286. doi: 10.1111/pce.13469

Bolger, A. M., Lohse, M., and Usadel, B. (2014). Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* 30, 2114–2120. doi: 10.1093/bioinformatics/btu170

Brehelin, C., Kessler, F., and van Wijk, K. J. (2007). Plastoglobules: versatile lipoprotein particles in plastids. *Trends Plant Sci.* 12, 260–266. doi: 10.1016/j.tplants.2007.04.003

Büchel, C. (2014). "FucoPxanthin-chlorophyll-proteins and non-photochemical fluorescence quenching of diatoms" in *Non-photochemical quenching and energy dissipation in plants, algae and cyanobacteria* 40, 259–275. doi: 10.1007/978-94-017-9032-1_11

Büchel, C. (2020). "Light-harvesting complexes of diatoms: Fucoxanthin-chlorophyll proteins" in *Photosynthesis in Algae: Biochemical and Physiological Mechanisms* 45, 441–457. doi: 10.1007/978-3-030-33397-3_16

Buck, J. M., Sherman, J., Bartulos, C. R., Serif, M., Halder, M., Henkel, J., et al. (2019). Lhcx proteins provide photoprotection via thermal dissipation of absorbed light in the diatom Phaeodactylum tricornutum. *Nat. Commun.* 10:4167. doi: 10.1038/s41467-019-12043-6

Chen, C., Chen, H., Zhang, Y., Thomas, H. R., Frank, M. H., He, Y., et al. (2020). TBtools: an integrative toolkit developed for interactive analyses of big biological data. *Mol. Plant* 13, 1194–1202. doi: 10.1016/j.molp.2020.06.009

Damiani, M. C., Popovich, C. A., Constenla, D., and Leonardi, P. I. (2010). Lipid analysis in Haematococcuspluvialis to assess its potential use as a biodiesel feedstock. *Bioresour. Technol.* 101, 3801–3807. doi: 10.1016/j.biortech.2009.12.136

Davidi, L., Levin, Y., Ben-Dor, S., and Pick, U. (2015). Proteome analysis of cytoplasmatic and plastidic beta-carotene lipid droplets in Dunaliella bardawil. *Plant Physiol.* 167, 60–79. doi: 10.1104/pp.114.248450

Deruere, J., Bouvier, F., Steppuhn, J., Klein, A., Camara, B., and Kuntz, M. (1994a). Structure and expression of two plant genes encoding chromoplast-specific proteins: occurrence of partially spliced transcripts. *Biochem. Biophys. Res. Commun.* 199, 1144–1150. doi: 10.1006/bbrc.1994.1350

Deruere, J., Romer, S., d'Harlingue, A., Backhaus, R. A., Kuntz, M., and Camara, B. (1994b). Fibril assembly and carotenoid overaccumulation in chromoplasts: a model for supramolecular lipoprotein structures. *Plant Cell* 6, 119–133. doi: 10.1105/tpc.6.1.119

Ding, X. T., Fan, Y., Jiang, E. Y., Shi, X. Y., Krautter, E., Hu, G. R., et al. (2021). Expression of the Vitreoscilla hemoglobin gene in Nannochloropsis oceanica regulates intracellular oxygen balance under high-light. *J. Photochem. Photobiol. B* 221:112237. doi: 10.1016/j.jphotobiol.2021.112237

Dodson, V. J., Mouget, J. L., Dahmen, J. L., and Leblond, J. D. (2014). The long and short of it: temperature-dependent modifications of fatty acid chain length and unsaturation in the galactolipid profiles of the diatoms Haslea ostrearia and Phaeodactylum tricornutum. *Hydrobiologia* 727, 95–107. doi: 10.1007/s10750-013-1790-4

Falkowski, P. G., Katz, M. E., Knoll, A. H., Quigg, A., Raven, J. A., Schofield, O., et al. (2004). The evolution of modern eukaryotic phytoplankton. *Science* 305, 354–360. doi: 10.1126/science.1095964

Field, C. B., Behrenfeld, M. J., Randerson, J. T., and Falkowski, P. (1998). Primary production of the biosphere: integrating terrestrial and oceanic components. *Science* 281, 237–240. doi: 10.1126/science.281.5374.237

Galas, L., Burel, C., Schapman, D., Ropitaux, M., Bernard, S., Benard, M., et al. (2021). Comparative structural and functional analyses of the fusiform, oval, and triradiate morphotypes of Phaeodactylum tricornutum Pt3 strain. *Front. Plant Sci.* 12:638181. doi: 10.3389/fpls.2021.638181

Goss, R., and Latowski, D. (2020). Lipid dependence of xanthophyll cycling in higher plants and algae. *Front. Plant Sci.* 11:455. doi: 10.3389/fpls.2020.00455

Gundermann, K., and Büchel, C. (2012). Factors determining the fluorescence yield of fucoxanthin-chlorophyll complexes (FCP) involved in non-photochemical quenching in diatoms. *Biochim. Biophys. Acta* 1817, 1044–1052. doi: 10.1016/j.bbabio.2012.03.008

Hu, H., and Pan, Y. (2020). Electroporation transformation protocol for Phaeodactylum tricornutum. *Methods Mol. Biol.* 2050, 163–167. doi: 10.1007/978-1-4939-9740-4_17

Jumper, J., Evans, R., Pritzel, A., Green, T., Figurnov, M., Ronneberger, O., et al. (2021). Highly accurate protein structure prediction with AlphaFold. *Nature* 596, 583–589. doi: 10.1038/s41586-021-03819-2

Kaup, M. T., Froese, C. D., and Thompson, J. E. (2002). A role for diacylglycerol acyltransferase during leaf senescence. *Plant Physiol.* 129, 1616–1626. doi: 10.1104/pp.003087

Kim, D., Langmead, B., and Salzberg, S. L. (2015). HISAT: a fast spliced aligner with low memory requirements. *Nat. Methods* 12, 357–360. doi: 10.1038/nmeth.3317

Kim, E. H., Lee, Y., and Kim, H. U. (2015). Fibrillin 5 is essential for plastoquinone-9 biosynthesis by binding to solanesyl diphosphate synthases in Arabidopsis. *Plant Cell* 27, 2956–2971. doi: 10.1105/tpc.15.00707

Kim, I., Lee, S. C., Kim, E. H., Song, K., Yang, T. J., and Kim, H. U. (2018). Genome-wide identification and expression analyses of the fibrillin family genes suggest their involvement in photoprotection in cucumber. *Plants-Basel* 7, 50–63. doi: 10.3390/plants7030050

Kuczynska, P., Jemiola-Rzeminska, M., Nowicka, B., Jakubowska, A., Strzalka, W., Burda, K., et al. (2020). The xanthophyll cycle in diatom Phaeodactylum tricornutum in response to light stress. *Plant Physiol. Biochem.* 152, 125–137. doi: 10.1016/j.plaphy.2020.04.043

Laizet, Y., Pontier, D., Mache, R., and Kuntz, M. (2004). Subfamily organization and phylogenetic origin of genes encoding plastid lipid-associated proteins of the fibrillin type. *J. Genome Sci. Technol.* 3, 19–28. doi: 10.1166/gl.2004.038

Latowski, D., Akerlund, H. E., and Strzalka, K. (2004). Violaxanthin de-epoxidase, the xanthophyll cycle enzyme, requires lipid inverted hexagonal structures for its activity. *Biochemistry* 43, 4417–4420. doi: 10.1021/bi049652g

Le Costaouëc, T., Unamunzaga, C., Mantecon, L., and Helbert, W. (2017). New structural insights into the cell-wall polysaccharide of the diatom Phaeodactylum tricornutum. *Algal Res.* 26, 172–179. doi: 10.1016/j.algal.2017.07.021

Lepetit, B., Goss, R., Jakob, T., and Wilhelm, C. (2012). Molecular dynamics of the diatom thylakoid membrane under different light conditions. *Photosynth. Res.* 111, 245–257. doi: 10.1007/s11120-011-9633-5

Letunic, I., Khedkar, S., and Bork, P. (2021). SMART: recent updates, new developments and status in 2020. *Nucleic Acids Res.* 49, D458–D460. doi: 10.1093/nar/gkaa937

Lichtenthaler, H. K. (1987). "Chlorophylls and carotenoids: pigments of photosynthetic biomembranes" in *Plant Cell Membranes*, eds. L. Packere and R. Douce, 350–382.

Lorente-Cebrian, S., Costa, A. G., Navas-Carretero, S., Zabala, M., Laiglesia, L. M., Martinez, J. A., et al. (2015). An update on the role of omega-3 fatty acids on inflammatory and degenerative diseases. *J. Physiol. Biochem.* 71, 341–349. doi: 10.1007/s13105-015-0395-y

Lundquist, P. K., Davis, J. I., and van Wijk, K. J. (2012a). ABC1K atypical kinases in plants: filling the organellar kinase void. *Trends Plant Sci.* 17, 546–555. doi: 10.1016/j.tplants.2012.05.010

Lundquist, P. K., Poliakov, A., Bhuiyan, N. H., Zybailov, B., Sun, Q., and van Wijk, K. J. (2012b). The functional network of the Arabidopsis plastoglobule proteome based on quantitative proteomics and genome-wide coexpression analysis. *Plant Physiol.* 158, 1172–1192. doi: 10.1104/pp.111.193144

Lundquist, P. K., Poliakov, A., Giacomelli, L., Friso, G., Appel, M., McQuinn, R. P., et al. (2013). Loss of plastoglobule kinases ABC1K1 and ABC1K3 causes conditional degreening, modified prenyl-lipids, and recruitment of the jasmonic acid pathway. *Plant Cell* 25, 1818–1839. doi: 10.1105/tpc.113.111120

Maeda, H., Hosokawa, M., Sashima, T., Funayama, K., and Miyashita, K. (2007). Effect of medium-chain triacylglycerols on anti-obesity effect of fucoxanthin. *J. Oleo Sci.* 56, 615–621. doi: 10.5650/jos.56.615

Martin, L. J. (2015). Fucoxanthin and its metabolite fucoxanthinol in cancer prevention and treatment. *Mar. Drugs* 13, 4784–4798. doi: 10.3390/md13084784

Ovide, C., Kiefer-Meyer, M. C., Berard, C., Vergne, N., Lecroq, T., Plasson, C., et al. (2018). Comparative in depth RNA sequencing of P. tricornutum's morphotypes reveals specific features of the oval morphotype. *Sci. Rep.* 8:14340. doi: 10.1038/s41598-018-32519-7

Pi, X., Zhao, S., Wang, W., Liu, D., Xu, C., Han, G., et al. (2019). The pigment-protein network of a diatom photosystem II-light-harvesting antenna supercomplex. *Science* 365, 463–474. doi: 10.1126/science.aax4406

Rey, P., Gillet, B., Romer, S., Eymery, F., Massimino, J., Peltier, G., et al. (2000). Over-expression of a pepper plastid lipid-associated protein in tobacco leads to changes in plastid ultrastructure and plant development upon stress. *Plant J.* 21, 483–494. doi: 10.1046/j.1365-313x.2000.00699.x

Rokkaku, T., Kimura, R., Ishikawa, C., Yasumoto, T., Senba, M., Kanaya, F., et al. (2013). Anticancer effects of marine carotenoids, fucoxanthin and its deacetylated product, fucoxanthinol, on osteosarcoma. *Int. J. Oncol.* 43, 1176–1186. doi: 10.3892/ijo.2013.2019

Schelbert, S., Aubry, S., Burla, B., Agne, B., Kessler, F., Krupinska, K., et al. (2009). Pheophytin pheophorbide hydrolase (pheophytinase) is involved in chlorophyll breakdown during leaf senescence in Arabidopsis. *Plant Cell* 21, 767–785. doi: 10.1105/tpc.108.064089

Singh, D. K., Maximova, S. N., Jensen, P. J., Lehman, B. L., Ngugi, H. K., and McNellis, T. W. (2010). FIBRILLIN4 is required for plastoglobule development and stress resistance in apple and Arabidopsis. *Plant Physiol.* 154, 1281–1293. doi: 10.1104/pp.110.164095

Singh, D. K., and McNellis, T. W. (2011). Fibrillin protein function: the tip of the iceberg? *Trends Plant Sci.* 16, 432–441. doi: 10.1016/j.tplants.2011.03.014

Souza, P. R., and Norling, L. V. (2016). Implications for eicosapentaenoic acid- and docosahexaenoic acid-derived resolvins as therapeutics for arthritis. *Eur. J. Pharmacol.* 785, 165–173. doi: 10.1016/j.ejphar.2015.05.072

Sun, Y., Xin, Y., Zhang, L., Wang, Y., Liu, R., Li, X., et al. (2022). Enhancement of violaxanthin accumulation in Nannochloropsis oceanica by overexpressing a carotenoid isomerase gene from Phaeodactylum tricornutum. *Front. Microbiol.* 13:942883. doi: 10.3389/fmicb.2022.942883

Tamura, K., Stecher, G., and Kumar, S. (2021). MEGA11: molecular evolutionary genetics analysis version 11. *Mol. Biol. Evol.* 38, 3022–3027. doi: 10.1093/molbev/msab120

Trapnell, C., Williams, B. A., Pertea, G., Mortazavi, A., Kwan, G., van Baren, M. J., et al. (2010). Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation. *Nat. Biotechnol.* 28, 511–515. doi: 10.1038/nbt.1621

van Wijk, K. J., and Kessler, F. (2017). Plastoglobuli: plastid microcompartments with integrated functions in metabolism, plastid developmental transitions, and environmental adaptation. *Annu. Rev. Plant Biol.* 68, 253–289. doi: 10.1146/annurev-arplant-043015-111737

Vartanian, M., Descles, J., Quinet, M., Douady, S., and Lopez, P. J. (2009). Plasticity and robustness of pattern formation in the model diatom Phaeodactylum tricornutum. *New Phytol.* 182, 429–442. doi: 10.1111/j.1469-8137.2009.02769.x

Vishnevetsky, M., Ovadis, M., and Vainstein, A. (1999). Carotenoid sequestration in plants: the role of carotenoid-associated proteins. *Trends Plant Sci.* 4, 232–235. doi: 10.1016/s1360-1385(99)01414-4

Wang, L. J., Fan, Y., Parsons, R. L., Hu, G. R., Zhang, P. Y., and Li, F. L. (2018). A rapid method for the determination of fucoxanthin in diatom. *Mar. Drugs* 16, 33–46. doi: 10.3390/md16010033

Wang, H., Wang, S., Chang, X., Hao, C., Sun, D., and Jing, R. (2019). Identification of TaPPH-7A haplotypes and development of a molecular marker associated with important agronomic traits in common wheat. *BMC Plant Biol.* 19:296. doi: 10.1186/s12870-019-1901-0

Wang, W., Yu, L. J., Xu, C., Tomizaki, T., Zhao, S., Umena, Y., et al. (2019). Structural basis for blue-green light harvesting and energy dissipation in diatoms. *Science* 363, 598–607. doi: 10.1126/science.aav0365

Wang, W., Zhao, S., Pi, X., Kuang, T., Sui, S. F., and Shen, J. R. (2020). Structural features of the diatom photosystem II-light-harvesting antenna complex. *FEBS J.* 287, 2191–2200. doi: 10.1111/febs.15183

Wolken, J. J., and Palade, G. E. (1952). Fine structure of chloroplasts in two flagellates. *Nature* 170, 114–115. doi: 10.1038/170114a0

Youssef, A., Laizet, Y., Block, M. A., Marechal, E., Alcaraz, J. P., Larson, T. R., et al. (2010). Plant lipid-associated fibrillin proteins condition jasmonate production under photosynthetic stress. *Plant J.* 61, 436–445. doi: 10.1111/j.1365-313X.2009.04067.x

Ytterberg, A. J., Peltier, J. B., and van Wijk, K. J. (2006). Protein profiling of plastoglobules in chloroplasts and chromoplasts. A surprising site for differential accumulation of metabolic enzymes. *Plant Physiol.* 140, 984–997. doi: 10.1104/pp.105.076083

Zhou, Q., Li, Q., Li, P., Zhang, S., Liu, C., Jin, J., et al. (2019). Carotenoid cleavage dioxygenases: identification, expression, and evolutionary analysis of this gene family in tobacco. *Int. J. Mol. Sci.* 20, 5796–5820. doi: 10.3390/ijms20225796

frontiers | Frontiers in Microbiology

TYPE Original Research
PUBLISHED 18 May 2023
DOI 10.3389/fmicb.2023.1124274

OPEN ACCESS

EDITED BY
Martin Hagemann,
University of Rostock, Germany

REVIEWED BY
Hui Wang,
Chinese Academy of Sciences (CAS), China
Maria Sinetova,
Timiryazev Institute of Plant Physiology (RAS), Russia

*CORRESPONDENCE
Jianping Yu
✉ jianping.yu@nrel.gov

RECEIVED 14 December 2022
ACCEPTED 05 May 2023
PUBLISHED 18 May 2023

CITATION
Cantrell M, Cano M, Sebesta J, Paddock T, Xiong W, Chou KJ and Yu J (2023) Manipulation of glycogen and sucrose synthesis increases photosynthetic productivity in cyanobacteria. *Front. Microbiol.* 14:1124274. doi: 10.3389/fmicb.2023.1124274

Manipulation of glycogen and sucrose synthesis increases photosynthetic productivity in cyanobacteria

Michael Cantrell, Melissa Cano, Jacob Sebesta, Troy Paddock, Wei Xiong, Katherine J. Chou and Jianping Yu*

Biosciences Center, National Renewable Energy Laboratory, Golden, CO, United States

Photosynthetic productivity is limited by low energy conversion efficiency in naturally evolved photosynthetic organisms, via multiple mechanisms that are not fully understood. Here we show evidence that extends recent findings that cyanobacteria use "futile" cycles in the synthesis and degradation of carbon compounds to dissipate ATP. Reduction of the glycogen cycle or the sucrose cycle in the model cyanobacterium *Synechocystis* 6803 led to redirection of cellular energy toward faster growth under simulated outdoor light conditions in photobioreactors that was accompanied by higher energy charge [concentration ratio of ATP/(ATP + ADP)]. Such manipulation of energy metabolism may have potential in engineering microalgal chassis cells to increase productivity of biomass or target metabolites.

KEYWORDS

photosynthesis, cyanobacteria, glycogen, sucrose, ATP

1. Introduction

Photosynthesis is nature's primary CO_2 capture and utilization (CCU) process, in which solar energy is captured by light harvesting pigments and stored in high-energy metabolites such as ATP and NADPH, which in turn drive CO_2 fixation and energy storage in carbon-carbon bonds. Production and consumption of high-energy metabolites dictate productivity, thus understanding their management is central to understanding photosynthetic energy conversion process, as well as the rational engineering of microalgal chassis cells for the synthesis of fuels and chemicals.

Since photosynthesis occurs in natural environments with dynamic light, temperature, and nutrient conditions, synthesis and consumption of ATP and NADPH must respond dynamically to changes in the environment. This drives the evolution of an abundance of mechanisms that help balance energy production with demands. This balance can be achieved by modulation of electron transport—either through dissipation by non-photochemical quenching or re-routing of electron transport via cyclic electron transport and alternative electron transport pathways (Allahverdiyeva et al., 2015). Biosynthetic pathways can also act as sinks for excess ATP and NADPH, and can serve to optimize metabolism in response to environmental conditions (Cano et al., 2018).

Synthesis of carbon reserves, such as glycogen in cyanobacteria as well as starch and lipids in eukaryotic algae, provides a sink for energy during nutrient limitation and serves as a reserve for cellular maintenance at night or during acclimation to changing light and

nutrient availability (Vitova et al., 2015; Luan et al., 2019; Ran et al., 2019). In the model cyanobacterium *Synechocystis* sp. PCC 6803 (hereafter referred to as *Synechocystis*), fixed carbon is committed to glycogen synthesis by ADP-glucose pyrophosphorylase (AGPase) encoded by the *glgC* gene (**Figure 1**; Luan et al., 2019). Previous characterization of a Δ*glgC* mutant that can no longer synthesize glycogen showed that the glycogen synthesis/degradation cycle acts as an energy management mechanism (Cano et al., 2018). Loss of this mechanism leads to a higher energy charge (EC; [ATP]/([ATP + ADP])) in the light, and to the excretion of organic acids under high light, mixotrophic conditions, or nitrogen deprivation (Carrieri et al., 2012; Grundel et al., 2012; Cano et al., 2018). Additionally, blocking glycogen synthesis could provide a means for increasing carbon partitioning into target molecules like mannitol and ethanol (Jacobsen and Frigaard, 2014; Namakoshi et al., 2016). However, a drawback of blocking glycogen synthesis is the reduced fitness under saturating and diurnal light conditions even in the presence of alternative sinks (Suzuki et al., 2010; Grundel et al., 2012).

Sucrose synthesis has been studied regarding its role in protection against increases in salinity. Aquatic environments face dynamic changes in salinity with depth and at mixing zones that drives osmotic loss of water, disrupting ion homeostasis and loss of cellular turgor pressure (Hagemann, 2011). To combat dynamic changes in salinity, cyanobacteria have evolved a suite of mechanisms that can rapidly adjust ion homeostasis. This involves transcriptional reprogramming, effectively increasing active export of Na^+ and Cl^-, and triggers the accumulation of osmo-protective organic compounds (Marin et al., 2004; Hagemann, 2011). In response to salt stress, *Synechocystis* accumulates primarily sucrose and glucosylglycerol (Kirsch et al., 2018, 2019). Salt shock is immediately followed by the rapid accumulation of sucrose within the first few hours, which is then followed by the slower accumulation of glucosylglycerol. After 24 h, sucrose content drops, and glucosylglycerol dominates as the primary osmolyte for long term accumulation in response to salt stress (Desplats et al., 2005). Sucrose accumulation is a product of salt ion activation of the sucrose phosphate synthetase (SPS; *sll0045*) and inactivation of a sucrose invertase (INV; *sll0626*) (Hagemann and Marin, 1999; Kirsch et al., 2018). Kirsch et al. (2018) demonstrated that the sucrose synthesis and degradation cycle is also active under non-stressed conditions, with Δ*inv* mutants displaying 10 times the sucrose accumulation as compared to wild-type. To our knowledge, sucrose synthesis has not been studied as an energy management mechanism.

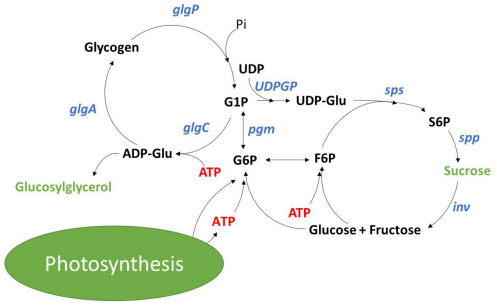

FIGURE 1
Glycogen and sucrose synthesis cycles in cyanobacteria with metabolites shown in black, enzymes shown in blue and osmolytes shown in green.

In this study, we show genetic and physiological evidence to support the hypothesis that cyanobacteria use the sucrose synthesis and degradation cycle in addition to the glycogen synthesis and degradation cycle to regulate ATP. In addition, we tested the hypothesis that a *partial inhibition* of glycogen synthesis would improve strain robustness while increasing cellular energy supply toward the synthesis of biomass. Disruption of such energy management mechanisms led to higher cellular energy levels and faster growth under a range of light conditions including simulated outdoor light conditions. These findings suggest a new approach to manipulate cellular energy metabolism toward higher photosynthetic productivity.

2. Results

2.1. Mutant verification

The *sps* gene in *Synechocystis* was completely knocked out by homologous recombination using a kanamycin resistance cassette (**Figure 2A**). The insertion was verified through PCR (**Figure 2B**). Complete removal of the *sps* coding sequence led to the loss of sucrose accumulation in exponential cultures acclimated to constant illuminations of 30, 200, and 500 μmol photons m^{-2} s^{-1} (**Figure 2C**) and after 12 h of salt stress (**Supplementary Figure 1**). A low glycogen strain (referred to as *LGS*) was generated as previously described in Carrieri et al. (2012). by complementation of a Δ*glgC* mutant, using the *glgC* coding sequence to replace the *psbA2* coding sequence (**Figures 2D–E**). This strain displayed reduced glycogen levels under constant illuminations of 30, 200, and 500 μmol photons m^{-2} s^{-1} (**Figure 2F**). This was correlated with lower sucrose concentrations in the *LGS* relative to WT under 30 and 200 μmol photons m^{-2} s^{-1} and after 12 h of incubation with 500 mM salt (**Figure 2C** and **Supplementary Figure 1**), indicating a link between the two biosynthetic pathways.

2.2. Sucrose or glycogen synthesis mutants exhibit increased energy charge and growth rates under a variety of light conditions

We hypothesized, based on prior observations on the Δ*glgC* mutant, that each strain may display higher ATP levels which may lead to faster growth under photoautotrophic conditions. The mutants were grown under constant light with 5% CO_2 in shake flask conditions, and under sinusoidal light in the NREL's lab-built Simulated Algal Growth Environment (SAGE) bioreactors, which are capable of supporting algal growth in 2-L bottles while mimicking outdoor light and temperature conditions (**Supplementary Figure 2A**). Intracellular ATP and ADP levels were assessed using a fast luciferase-based assay as described in Cano et al. (2018). Under a continuous light of 30 μmol photons m^{-2} s^{-1}, we observed increases in growth rate for Δ*sps* vs. wild-type (11.3%; **Table 1**) and found that both Δ*sps* and *LGS* displayed

10.3389/fmicb.2023.1124274

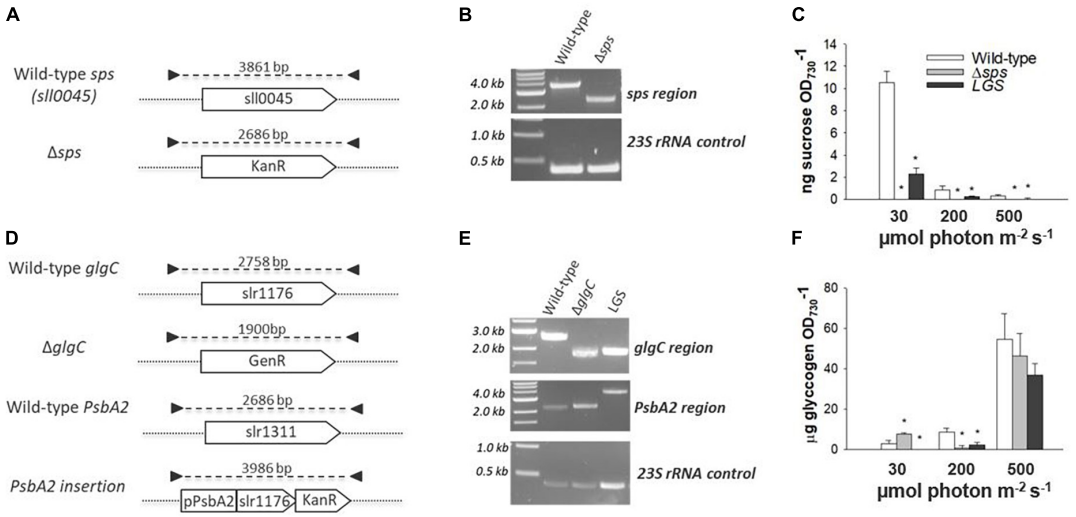

FIGURE 2
Mutant generation, genotyping by PCR, and phenotype verification. The Δ*sps* mutant was generated by complete removal of the *sps* coding region via homologous recombination with a kanamycin resistance cassette [*KanR*; **(A)**] and homoplasmy was verified by PCR **(B)**. Sucrose content was assayed during exponential growth in cultures acclimated to 30, 200, and 500 μmol photons m^{-2} s^{-1} **(C)**. The low glycogen strain (*LGS*) strain was generated in the Δ*glgC* background (Gentamycin Resistant—*GenR*) by insertion of a *glgC* coding sequence with kanamycin resistance (*KanR*) under the regulation of the native PsbA2 promoter **(D,E)**. Glycogen content was assayed during exponential growth in cultures acclimated to constant 30, 200, and 500 μmol photons m^{-2} s^{-1} **(F)**. The 23S ribosomal RNA was used as an amplification control for each PCR reaction. Symbols (*) represent significant differences from wild-type within light acclimated states based on a one-way ANOVA and Tukey's HSD *post-hoc* test ($p < 0.05$).

significant increases in energy charge ([ATP]/[ATP + ADP]; Figure 3D). Under a continuous 200 μmol photons m^{-2} s^{-1} Δ*sps* and *LGS* exponential growth rates were 18.7% ($p < 0.05$) and 8.7% faster than wild-type (Figure 3B and Table 1), and had significantly higher energy charge (Figure 3E). Under 500 μmol photons m^{-2} s^{-1}, each strain exhibited reduced growth rates relative to 200 μmol photons m^{-2} s^{-1} and energy charge was significantly increased only during early stationary phase in mutants (Figure 3F). Under these saturating light conditions, only the *LGS* displayed significantly faster growth than wild-type (6.8%, $p < 0.05$).

TABLE 1 Effects of acclimation state on mutant exponential growth rate and pigment content for shake flask cultures grown under constant lights intensities of 30, 200, and 500 μmol photons m^{-2} s^{-1}.

30 μmol photons m^{-2} s^{-1}				
	Units	Wild-type	Δ*sps*	*LGS*
Growth rate	Day^{-1}	0.53 ± 0.02	0.59 ± 0.03*	0.54 ± 0.01
Chlorophyll *a*	μg OD^{-1}	1.00 ± 0.06	0.64 ± 0.02*	0.79 ± 0.12*
Carotenoids	μg OD^{-1}	0.31 ± 0.03	0.31 ± 0.01	0.32 ± 0.02
200 μmol photons m^{-2} s^{-1}				
	Units	Wild-type	Δ*sps*	*LGS*
Growth rate	Day^{-1}	2.68 ± 0.22	3.18 ± 0.03*	2.91 ± 0.01
Chlorophyll *a*	μg OD^{-1}	1.01 ± 0.20	0.65 ± 0.17	0.74 ± 0.01
Carotenoids	μg OD^{-1}	0.29 ± 0.02	0.31 ± 0.03	0.34 ± 0.03
500 μmol photons m^{-2} s^{-1}				
	Units	Wild-type	Δ*sps*	*LGS*
Growth rate	Day^{-1}	2.50 ± 0.03	2.62 ± 0.07	2.67 ± 0.04*
Chlorophyll *a*	μg OD^{-1}	0.45 ± 0.03	0.45 ± 0.04	0.5 ± 0.08
Carotenoids	μg OD^{-1}	0.40 ± 0.02	0.43 ± 0.06	0.38 ± 0.06

Data represents the mean ± s.d. (n = 3–4). Symbols (*) represent significant differences from wild-type within light acclimated states based on a one-way ANOVA and Tukey's HSD *post-hoc* test ($p < 0.05$).

2.3. Increased growth rates are correlated with high photosynthetic rates during growth in moderate/high light

We investigated the impact of each genetic modification on photosynthesis by measuring photosynthesis vs. irradiances (PI curve) with sequential 60 s light steps supplemented with 10 mM sodium bicarbonate. Across the constant light conditions investigated, we found that the maximum photosynthetic rate (Pmax; Figure 4) and saturation point of photosynthesis (IK) increased with culture light intensity (Supplementary Figure 3). When comparing between strains, we found that only mutants grown under 200 μmol photons m^{-2} s^{-1} displayed a significant increase in Pmax (Figures 4B, D). Under these conditions both Δ*sps* and *LGS* displayed a 61–69% increase in Pmax with no significant differences observed between light limited slope (α) and IK (e).

2.4. Mutants show increases in the accumulation rates of ash free dry weight (AFDW) in SAGE reactors and a higher energy charge at dawn

To determine if this improved growth is conserved under outdoor light conditions, we investigated productivity based

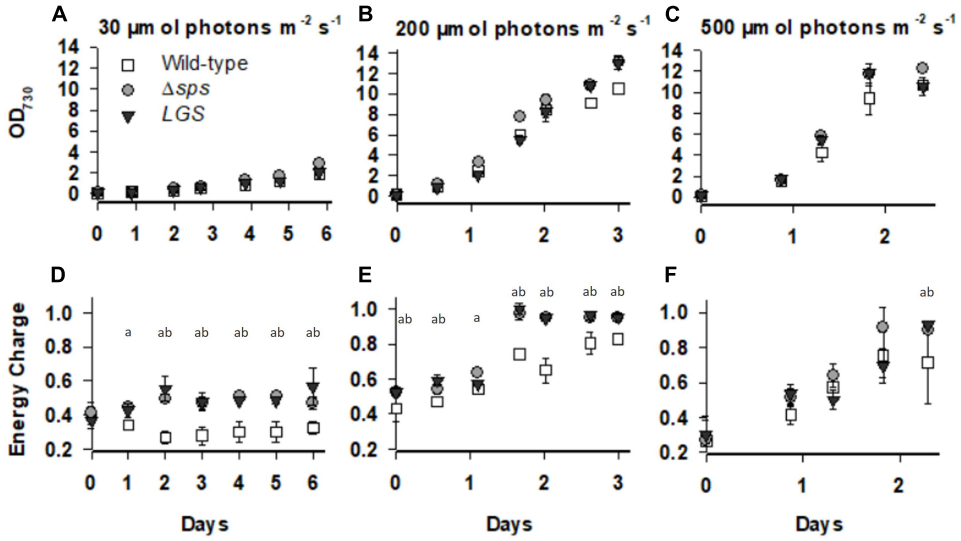

FIGURE 3
The optical density and energy charge ([ATP]/([ATP] + [ADP])) for wild-type (open squares), Δ*sps* (shaded circles), and *LGS* (black triangles) grown under constant lights intensities of 30 **(A,D)**, 200 **(B,E)**, and 500 **(C,F)** µmol photons m^{-2} s^{-1}. Data shown reflects n = 3–5 cultures with error bars reflecting the standard deviation. Letters denote significant differences based on a one-way ANOVA ($P < 0.05$) and Tukey's HSD *post-hoc* test with (a) denoting significant differences between wild-type and Δ*sps* and (b) denoting significant differences between wild-type and *LGS*.

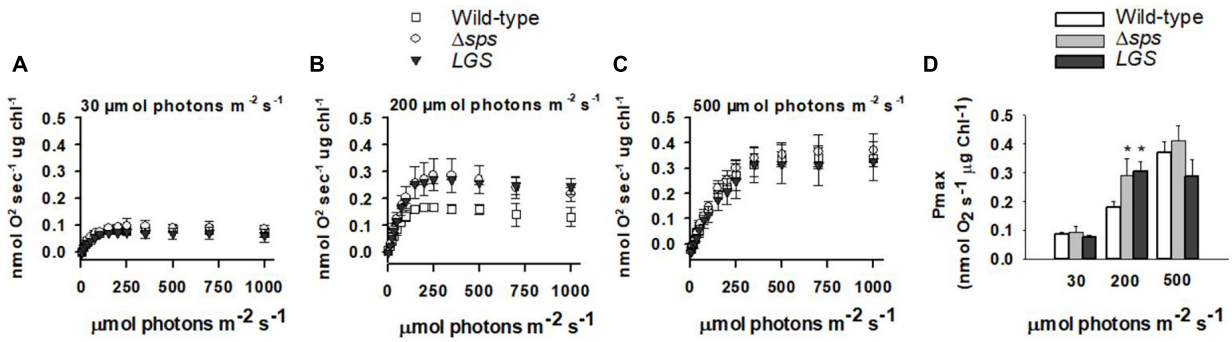

FIGURE 4
Photosynthesis vs. irradiance for wild-type (open squares), Δ*sps* (shaded circles), and *LGS* (black triangles) grown under constant lights intensities of 30 **(A)**, 200 **(B)**, and 500 µmol photons m^{-2} s^{-1} **(C)** and maximum photosynthetic rate (Pmax) for each culture condition **(D)**. Values displayed reflect oxygen evolution rates (nmol O_2 s^{-1} µg chl^{-1}) for samples exposed to sequentially increasing light intensities. Data shown reflects n = 3 with error bars reflecting the standard deviation. Symbols (*) represent significant differences from wild-type within light acclimated states based on a one-way ANOVA and Tukey's HSD *post-hoc* test ($p < 0.05$).

on AFDW accumulation in the NREL simulated algae growth environment (SAGE) reactors. These were operated at a constant 30 °C under light conditions simulating those found during the fall at the Arizona Center for Algae Technology and Innovation (AZCATI) test site, which followed a sinusoidal light rhythm with intensities peaking at 1,700 photons m^{-2} s^{-1} (**Supplementary Figures 2A, B**). AFDW was assessed daily at solar dawn for cultures after 3 days of pre-acclimation under this light regime. We found that Δ*sps* and *LGS* displayed increases in AFDW accumulation of 17 and 19%, respectively (**Figure 5**). In the first 3 days in particular, the sps mutant grew much faster with a biomass accumulation rate 29% higher than the wild type.

In addition to a high rate of AFDW accumulation, we found that under sinusoidal light conditions cultures maintained a significantly higher energy charge 30 min after dawn (at Zeitgeber time 0.5 h) (**Figure 6**).

3. Discussion

Understanding the regulation of high energy molecules such as ATP and NADPH is central in understanding biological processes such as photosynthesis, and in engineering microalgal chassis cells for carbon sequestration and utilization. This work expanded our prior findings on mechanisms for intracellular energy management in *Synechocystis*. Previous work demonstrated a high energy charge in a mutant that can no longer accumulate glycogen (Δ*glgC*). However, this strain had reduced fitness under saturating light and diurnal conditions (Grundel et al., 2012;

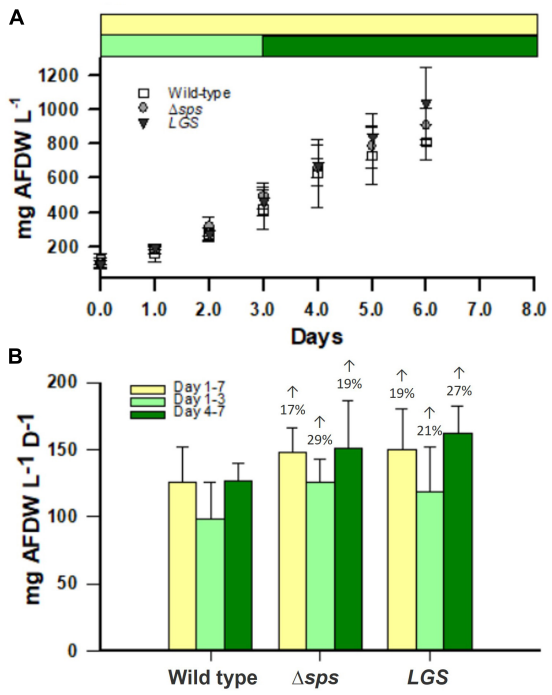

FIGURE 5
Ash free dry weight (AFDW) accumulation in SAGE reactors **(A)** for wild-type (open squares), Δ*sps* (shaded circles), and *LGS* (black triangles). Accumulation rates are shown for SAGE reactors **(B)** for the entire growth period (yellow), the beginning of the growth period (light green) and the end of the growth period (dark green). Data shown reflects $n = 4$ cultures with error bars reflecting the standard deviation. Percentages provided above bars in **(B)** denote significant differences from wild-type based on a one-way ANOVA and Tukey's HSD *post-hoc* test ($P < 0.05$).

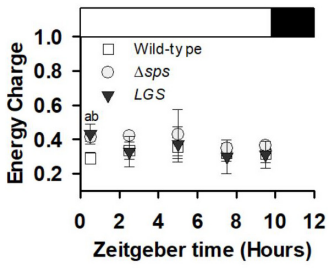

FIGURE 6
Energy charge ([ATP]/([ATP] + [ADP])) collected during early exponential growth (day 3) for cultures grown under sinusoidal light in SAGE reactors. Data shown reflects $n = 3–5$ with error bars reflecting the standard deviation. Letters denote significant differences based on a one-way ANOVA ($P < 0.05$) and Tukey's HSD *post hoc* test with (a) denoting significant differences between wild-type and Δsps and (b) denoting significant differences between wild-type and *LGS*.

Cano et al., 2018). Therefore, we hypothesized that a strain with low glycogen content may display high EC and improved photosynthetic rates without reducing fitness. Here we found that the complemented Δ*glgC* mutant strain [*LGS* (Carrieri et al., 2012)] displayed increased EC as well as growth rates.

Glycogen plays an essential role in both balancing energy homeostasis and initiation of photosynthetic metabolism. Shinde et al. (2020) demonstrated that glycogen degradation is required for initiating photosynthesis upon transition from dark to light conditions. Furthermore, they showed that absence of glycogen synthesis in the Δ*glgC* mutant or the oxidative pentose phosphate pathway in the Δ*gnd* mutant led to reduced photosynthetic efficiency and slower PSI ($P700^+$) re-reduction rates. This was used to support the hypothesis that glycogen degradation through the OPP pathway is essential for initiation of photosynthesis by generating the NADPH needed to activate key Calvin-Benson cycle enzymes (Shinde et al., 2020). We previously showed that the Δ*glgC* mutant under photoautotrophic conditions experiences metabolite overflow, with both increased EC and organic acid excretion to the extracellular medium (Carrieri et al., 2012, 2015; Cano et al., 2018). Taken together, these results indicate that glycogen synthesis and degradation facilitate metabolic homeostasis by either generating NADPH when reducing energy is low or can act as a sink for carbon and ATP when energy is in excess. Here we found that faster growth in our LGS was correlated with higher energy charge (**Figures 3**, **6**) and a lower glycogen content (**Figure 2F** and **Supplementary Figure 4**). In contrast to moderate and diurnal light levels, our LGS displayed comparable growth rates and glycogen levels to wild-type under constant saturating light intensities (500 μmol photon m^{-2} s^{-1}), reflecting light induction of the *psbA2* promoter used for complementation. This suggests that even under non-stress conditions, cyanobacteria use a glycogen synthesis/degradation "futile cycle" to help maintain energy balance and that reducing glycogen storage under non-saturating light conditions can enhance biomass productivity.

Sucrose synthesis is essential for adaption to salt stress. Previous characterization of the Δ*sps* mutant demonstrated its importance for growth under salt stress, with Δ*sps* displaying lower growth in BG-11 supplemented with 684 mM NaCl. In the wild-type, despite a consistent growth reduction, 684 mM NaCl leads to a transient sucrose accumulation that peaks at 4 h and then drops off significantly after 12 h (Desplats et al., 2005). This decrease of sucrose levels reflects the activity of invertase, with the *Synechocystis* invertase mutant displaying no drop off in sucrose under salt stress and ~9 times the sucrose concentration in standard BG-11 (Kirsch et al., 2018). Here we demonstrate that without inducing salt stress (growth in normal BG-11), the Δ*sps* mutant displays increased EC and improved growth under continuous moderate light and simulated outdoor light conditions (diurnal sinusoidal fluctuating light with high maximum intensity) (**Figures 3**, **5**, **6**). The higher EC in this case reflects the savings of two ATP per sucrose synthesis cycle and indicates that the sucrose cycle plays a role in energy regulation under constant light conditions. The observations of higher EC and faster growth in the Δ*sps* mutant indicate that faster growth is not just a consequence of carbon redirection, as there is minimal sucrose accumulation under these conditions.

Sucrose and glycogen synthesis reflect metabolically linked cycles and their reduction increases fitness under moderate light. Previous efforts to modify carbon partitioning in cyanobacteria have highlighted the interconnection of the sucrose and glycogen cycles. Sucrose and glycogen synthesis both involve glucose-6-phosphate generated from the Calvin Benson cycle or through

regeneration of glucose through glycogen or sucrose catabolism. Sucrose and glycogen pools appear linked, though flux between them is dependent on environmental conditions. Under salt stress, knockdown of glycogen synthesis (*glgC*) limited sucrose production in *Synechococcus* PCC7942, indicating that under environmental stress glycogen acts as a carbon pool for sucrose synthesis (Qiao et al., 2018). This is supported by the observations that overexpression of both *glgC* and *sps* can significantly increase sucrose accumulation, and by similar observations on the dependence of sucrose synthesis on the glycogen pool in *Synechococcus* UTEX 2973 under salt stress (Qiao et al., 2018; Lin et al., 2020). In contrast, under non-salt stress conditions, deletion of *glgC* in *Synechococcus* PCC7942 was shown to increase sucrose accumulation rates by 10–15% (Ducat et al., 2012). This suggests that under normal conditions, sucrose, glycogen synthesis and anabolic metabolism all compete for glucose-6-phosphate. We found that either knocking out sucrose synthesis or reducing glycogen synthesis led to similar phenotypes. While both appears to be impacted in their acclimation state based on their chlorophyll content under low and moderate constant light, each mutant displayed increases in growth rates, especially under diurnal conditions (**Figure 5**). These increases were correlated with a high EC, which may also facilitate improved photosynthetic function observed at constant irradiances of 200 μmol photons m^{-2} s^{-1} (**Figures 3**, **4**).

4. Materials and methods

4.1. Growth conditions

Wild-type *Synechocystis* sp. PCC6803, the *LGS*, and Δ*sps* strains were grown photoautotrophically in standard BG-11 medium supplemented with 100 mM $NaHCO_3$, 20 mM TES (pH 8) in 250 ml flasks unless otherwise noted. The *LGS* strain was constructed as previously described in Carrieri et al. (2012) and the Δ*sps* mutant was constructed by homologous recombination with the pIDTSMART vector containing a kanamycin cassette flanked by *sps* 5′ and 3′ regions. Segregation of the Δ*sps* mutant was verified by PCR product analysis. Δ*sps* and the *LGS* were maintained on plates of BG-11 with 25 μg ml^{-1} kanamycin or 25 μg ml^{-1} kanamycin and 25 μg ml^{-1} gentamycin, respectively. For shake flask experiments, liquid cultures were inoculated from plates and allowed to acclimate to the experimental conditions used for 3–6 days with dilution every 2–3 days to maintain cultures in exponential growth.

Shake flask cultures were grown in a Percival growth chamber maintained at 30°C with 5% ambient CO_2 under each respective light conditions. Moderate light intensities (30, 200, and 500 μmol photons m^{-2} s^{-1}) were provided by cool light fluorescent lamps. Culture growth and dilution was performed using absorbance measurements at 730 nm ($OD_{730\ nm}$) with a Beckman Coulter DU 900 spectrometer. All experiments used cultures inoculated at an $OD_{730\ nm}$ of 0.09–0.10 from an exponentially grown culture—tested to be axenic by plating on LB medium. Exponential growth rates were calculated by linear regression using the exponential growth equation $\mu = \frac{\ln(F)-\ln(I)}{\Delta t}$ where F is the final cell concentration, I the initial cell concentration, μ is the specific growth rate (day^{-1}) and t time in days.

SAGE reactor experiments were maintained at 30°C by water bath and illuminated using light panels (Renology, Ontario, CA, USA). We programmed this light panel to mimic daily light patterns recorded at the Arizona Center for Algae Technology and Innovation (AZCATI) during a week in September. This generally followed a 12:12 light: dark sinusoidal light cycle with peak light intensities of 1554 μmol photons m^{-2} s^{-1} (**Supplementary Figure 2B**).

4.2. ATP and ADP determination

ATP and ADP determination was performed as described in Cano et al. (2018) using the ATP/ADP ratio assay kit (Sigma Aldrich, cat. no. MAK135). In short, 10 μl culture was sampled from each flask directly into 96 well plate containing 90 μl sigma cell lysis reagent. Luminescence was measured as described by the manufacturer using a Tecan Infinite M200 Pro luminometer. Energy charge as shown throughout the manuscript reflects the concentration [ATP]/([ATP] + [ADP]).

4.3. Glycogen quantification

Glycogen quantification was performed as described in Vidal and Venegas-Caleron (2019). In short, 10 ml of sample (2–3 OD units) was collected by repeated centrifugation at 10,000 × *g* at room temperature (about 23°C) in a 2 ml Eppendorf tube, decanted and immediately frozen. Alkaline thermolysis was performed by boiling in 30% KOH for 90 min at 95°C in a dry bath. Glycogen was then precipitated from the raw extract with 2 ml of cold, 100% ethanol at −80°C overnight. Precipitated glycogen was then collected by centrifugation at 16,000 × *g* for 30 min at 4°C. Isolated glycogen was digested to glucose by incubation with 10 units of aminoglucosidase (Sigma Aldrich, cat. no. A7095) at 55°C for 2 h. Glucose quantification was then performed on dilutions using the glucose assay kit (Sigma Aldrich, cat. no. GAGO-20) (Vidal and Venegas-Caleron, 2019).

4.4. Sucrose quantification

Sucrose quantification was performed as described in Kirsch et al. (2019) with the following modifications. Samples were harvested by centrifugation at 10,000 × *g* at room temperature in a 2 ml Eppendorf tube, decanted and immediately frozen. Pelleted samples were resuspended in 80% ethanol and incubated overnight at 65°C. After centrifugation the supernatants were collected and dried by a speedvac concentrator (Savant ISS110). Dried samples were resuspended in diH_2O and used for sucrose quantification using a Sucrose quantification kit (Sigma Aldrich, cat. no. MAK267) following the manufacturer's instructions (Kirsch et al., 2019).

4.5. Dry weight and productivity determination

Dry weights were determined for 10 ml samples filtered onto pre-washed Whatman glass microfiber filters (Fisher cat. no. 18-250-70) in Aluminum Tins (Fisher cat. no. 08-732-102). Samples were initially dried in an oven at 55°C and then allowed to come to room temperature before being weighed to obtain the total suspended solids. Ash content of samples was subsequently determined by burning dry samples in a furnace at 575°C for 180 min. Samples were allowed to cool for 30 min prior to weight determination of the ash content of samples. AFDW was determined by subtracting the ash content from the total suspended solids and was normalized by volume for calculations of AFDW accumulation.

4.6. Chlorophyll and total carotenoid quantification

Pigments were extracted in 100% methanol overnight and spectra was collected using a Beckman coulter DU 900 spectrometer. Samples were pelleted by centrifugation at 17,000 × *g* for 10 min with 0.01% (v/v) Tween 20. Pellets generated were dissolved in methanol overnight at 4°C and centrifuged at 17,000 × *g* for 1 min to clarify solutions prior to spectra collection. Chl *a* and total carotenoid content were determined as previously described (Wellburn, 1994; Ritchie, 2006).

4.7. Photosynthesis vs. irradiance curves

Rapid light curves were generated using a Clark electrode and 1 min light steps. Samples were harvested and immediately transferred to the Clark electrode cuvette (3–6 μg Chlorophyll *a*/ml) with 10 mM sodium bicarbonate and dark incubated at 30°C for 10 min. After dark incubation, cells were exposed to sequential 1 min light levels of 5, 15, 25, 35, 50, 75, 100, 150, 200, 250, 350, 500, 700, and 1000 μmol photon m^{-2} s^{-1}. P vs. I curves were fit using the exponential difference equation using the curve fitting tool described by Ritchie (2008) to derive the maximum oxygen evolution rate (Pmax), the light limited slope (α) and the irradiance at saturation (IK) (Ritchie, 2008).

4.8. Statistics

All measurements were made on at least three independent cultures. Statistically significant differences were assessed by one-way ANOVA using Tukey's honestly significant differences test (Tukey's HSD) to delineate between significantly different groups.

Data availability statement

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

Author contributions

MiC and JY designed the study. MiC, TP, and MeC collected the data. MiC, WX, KC, and JY analyzed the data. MiC drafted the manuscript. All authors edited manuscript and agreed on the submitted manuscript.

Funding

This work was provided by the U.S. Department of Energy, Office of Energy Efficiency and Renewable Energy, Bioenergy Technologies Office (to JY). Part of the work was also supported by National Renewable Energy Laboratory LDRD Program.

Acknowledgments

This work was authored in part by the National Renewable Energy Laboratory, operated by Alliance for Sustainable Energy, LLC, for the U.S. Department of Energy (DOE) under Contract No. DE-AC36-08GO28308. Nick Sweeney helped with SAGE reactors.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

Author disclaimer

Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fmicb.2023.1124274/full#supplementary-material

SUPPLEMENTARY FIGURE 1
Sucrose accumulation in mutants after 12 h of salt stress under 30 μmol photons m^{-2} s^{-1}. Symbols (*) denote significant differences based on a one-way ANOVA ($P < 0.05$) from wild-type. Data shown reflects $n = 3$ with error bars reflecting the standard deviation.

SUPPLEMENTARY FIGURE 2
SAGE reactor culture conditions and OD_{730} collected for each experiment. **(A)** Image of SAGE reactor culture conditions. **(B)** Light regime provided to SAGE reactor cultures based on irradiances collected at the AZCATI cultivation site. **(C)** OD_{730} measurements for culture growth with 3 days of pre-acclimation.

SUPPLEMENTARY FIGURE 3
Light limited slope (α) and irradiance at saturation (IK) for P vs. I curves under different constant light culture conditions. Data shown reflects $n = 3–5$ cultures with error bars reflecting the standard deviation.

SUPPLEMENTARY FIGURE 4
Glycogen quantification for SAGE reactor cultures. Data shown reflects $n = 3–5$ cultures with error bars reflecting the standard deviation. Letters denote significant differences based on a one-way ANOVA ($P < 0.05$) and Tukey's HSD *post-hoc* test with (a) denoting significant differences between wild-type and Δ*sps* and (b) denoting significant differences between wild-type and *LGS*.

References

Allahverdiyeva, Y., Suorsa, M., Tikkanen, M., and Aro, E. M. (2015). Photoprotection of photosystems in fluctuating light intensities. *J. Exp. Bot.* 66, 2427–2436. doi: 10.1093/jxb/eru463

Cano, M., Holland, S. C., Artier, J., Burnap, R. L., Ghirardi, M., Morgan, J. A., et al. (2018). Glycogen synthesis and metabolite overflow contribute to energy balancing in *Cyanobacteria*. *Cell Rep.* 23, 667–672. doi: 10.1016/j.celrep.2018.03.083

Carrieri, D., Broadbent, C., Carruth, D., Paddock, T., Ungerer, J., Maness, P. C., et al. (2015). Enhancing photo-catalytic production of organic acids in the cyanobacterium *Synechocystis* sp.Pcc 6803 ΔglgC, a strain incapable of glycogen storage. *Microb. Biotechnol.* 8, 275–280. doi: 10.1111/1751-7915.12243

Carrieri, D., Paddock, T., Maness, P.-C., Seibert, M., and Yu, J. (2012). Photo-catalytic conversion of carbon dioxide to organic acids by a recombinant cyanobacterium incapable of glycogen storage. *Energy Environ. Sci.* 5, 9457–9461. doi: 10.1039/c2ee23181f

Desplats, P., Folco, E., and Salerno, G. L. (2005). Sucrose may play an additional role to that of an osmolyte in *Synechocystis* sp. Pcc 6803 salt-shocked cells. *Plant Physiol. Biochem.* 43, 133–138. doi: 10.1016/j.plaphy.2005.01.008

Ducat, D. C., Avelar-Rivas, J. A., Way, J. C., and Silvera, P. A. (2012). Rerouting carbon flux to enhance photosynthetic productivity. *Appl. Environ. Microbiol.* 78, 2660–2668. doi: 10.1128/AEM.07901-11

Grundel, M., Scheunemann, R., Lockau, W., and Zilliges, Y. (2012). Impaired glycogen synthesis causes metabolic overflow reactions and affects stress responses in the cyanobacterium *Synechocystis* sp. Pcc 6803. *Microbiology* 158, 3032–3043. doi: 10.1099/mic.0.062950-0

Hagemann, M. (2011). Molecular biology of cyanobacterial salt acclimation. *FEMS Microbiol. Rev.* 35, 87–123. doi: 10.1111/j.1574-6976.2010.00234.x

Hagemann, M., and Marin, K. (1999). Salt-induced sucrose accumulation is mediated by sucrose-phosphate-synthase in *Cyanobacteria*. *J. Plant Physiol.* 155, 424–430. doi: 10.1016/S0176-1617(99)80126-6

Jacobsen, J. H., and Frigaard, N. U. (2014). Engineering of photosynthetic mannitol biosynthesis from Co2 in a cyanobacterium. *Metab. Eng.* 21, 60–70. doi: 10.1016/j.ymben.2013.11.004

Kirsch, F., Klahn, S., and Hagemann, M. (2019). Salt-regulated accumulation of the compatible solutes sucrose and glucosylglycerol in *Cyanobacteria* and Its biotechnological potential. *Front. Microbiol.* 10:2139. doi: 10.3389/fmicb.2019.02139

Kirsch, F., Luo, Q., Lu, X., and Hagemann, M. (2018). Inactivation of invertase enhances sucrose production in the cyanobacterium *Synechocystis* sp. Pcc 6803. *Microbiology* 164, 1220–1228. doi: 10.1099/mic.0.000708

Lin, P. C., Zhang, F., and Pakrasi, H. B. (2020). Enhanced production of sucrose in the fast-growing cyanobacterium *Synechococcus elongatus* Utex 2973. *Sci. Rep.* 10:390. doi: 10.1038/s41598-019-57319-5

Luan, G., Zhang, S., Wang, M., and Lu, X. (2019). Progress and perspective on cyanobacterial glycogen metabolism engineering. *Biotechnol. Adv.* 37, 771–786. doi: 10.1016/j.biotechadv.2019.04.005

Marin, K., Kanesaki, Y., Los, D. A., Murata, N., Suzuki, I., and Hagemann, M. (2004). Gene expression profiling reflects physiological processes in salt acclimation of *Synechocystis* sp. strain Pcc 6803. *Plant Physiol.* 136, 3290–3300. doi: 10.1104/pp.104.045047

Namakoshi, K., Nakajima, T., Yoshikawa, K., Toya, Y., and Shimizu, H. (2016). Combinatorial deletions of glgC and phace enhance ethanol production in *Synechocystis* sp. Pcc 6803. *J. Biotechnol.* 239, 13–19. doi: 10.1016/j.jbiotec.2016.09.016

Qiao, C., Duan, Y., Zhang, M., Hagemann, M., Luo, Q., and Lu, X. (2018). Effects of reduced and enhanced glycogen pools on salt-induced sucrose production in a sucrose-secreting strain of *Synechococcus elongatus* Pcc 7942. *Appl. Environ. Microbiol.* 84:e02023-17. doi: 10.1128/AEM.02023-17

Ran, W., Wang, H., Liu, Y., Qi, M., Xiang, Q., Yao, C., et al. (2019). Storage of starch and lipids in microalgae: biosynthesis and manipulation by nutrients. *Bioresour. Technol.* 291:121894. doi: 10.1016/j.biortech.2019.121894

Ritchie, R. J. (2006). Consistent sets of spectrophotometric chlorophyll equations for acetone, methanol and ethanol solvents. *Photosynth Res.* 89, 27–41. doi: 10.1007/s11120-006-9065-9

Ritchie, R. J. (2008). Fitting light saturation curves measured using modulated fluorometry. *Photosynth Res.* 96, 201–215. doi: 10.1007/s11120-008-9300-7

Shinde, S., Zhang, X., Singapuri, S. P., Kalra, I., Liu, X., Morgan-Kiss, R. M., et al. (2020). Glycogen metabolism supports photosynthesis start through the oxidative pentose phosphate pathway in *Cyanobacteria*. *Plant Physiol.* 182, 507–517. doi: 10.1104/pp.19.01184

Suzuki, E., Ohkawa, H., Moriya, K., Matsubara, T., Nagaike, Y., Iwasaki, I., et al. (2010). Carbohydrate metabolism in mutants of the cyanobacterium *Synechococcus elongatus* Pcc 7942 defective in glycogen synthesis. *Appl. Environ. Microbiol.* 76, 3153–3159. doi: 10.1128/AEM.00397-08

Vidal, R., and Venegas-Caleron, M. (2019). Simple, fast and accurate method for the determination of glycogen in the model unicellular cyanobacterium *Synechocystis* sp. Pcc 6803. *J. Microbiol. Methods* 164:105686. doi: 10.1016/j.mimet.2019.105686

Vitova, M., Bisova, K., Kawano, S., and Zachleder, V. (2015). Accumulation of energy reserves in algae: from cell cycles to biotechnological applications. *Biotechnol. Adv.* 33, 1204–1218. doi: 10.1016/j.biotechadv.2015.04.012

Wellburn, A. R. (1994). The spectral determination of chlorophylls a and b, as well as Total carotenoids, using various solvents with spectrophotometers of different resolution. *J. Plant Physiol.* 144, 307–313. doi: 10.1016/S0176-1617(11)81192-2

Frontiers in Microbiology

Explores the habitable world and the potential of microbial life

The largest and most cited microbiology journal which advances our understanding of the role microbes play in addressing global challenges such as healthcare, food security, and climate change.

Discover the latest Research Topics

See more →

Frontiers

Avenue du Tribunal-Fédéral 34
1005 Lausanne, Switzerland
frontiersin.org

Contact us

+41 (0)21 510 17 00
frontiersin.org/about/contact