Genomic analysis and biochemical profiling of an unaxenic strain of *Synechococcus* sp. isolated from the Peruvian Amazon Basin region

Marianela Cobos1,2,3, Ruth C. Condori1, Miguel A. Grandez4, Segundo L. Estela1, Marjorie T. Del Aguila1,2, Carlos G. Castro1,2, Hicler N. Rodríguez2, Jhon A. Vargas2,4, Alvaro B. Tresierra1, Luis A. Barriga1, Jorge L. Marapara2,3, Pedro M. Adranzén2,3, Roger Ruiz5 and Juan C. Castro2,3*

1Laboratorio de Biotecnología y Bioenergética (LBB), Universidad Científica del Perú (UCP), Iquitos, Perú, 2Unidad Especializada de Biotecnología, Centro de Investigación de Recursos Naturales de la UNAP (CIRNA), Universidad Nacional de la Amazonia Peruana (UNAP), Iquitos, Perú, 3Departamento Académico de Ciencias Biomédicas y Biotecnología, Facultad de Ciencias Biológicas, Universidad Nacional de la Amazonia Peruana (UNAP), Ciudad Universitaria de Zungarococha, Iquitos, Perú, 4Instituto de Física de São Carlos, Universidade de São Paulo, São Carlos, Brazil, 5Facultad de Industrias Alimentarias, Universidad Nacional de la Amazonia Peruana (UNAP), Ciudad Universitaria de Zungarococha, Iquitos, Perú

Cyanobacteria are diverse photosynthetic microorganisms able to produce a myriad of bioactive chemicals. To make possible the rational exploitation of these microorganisms, it is fundamental to know their metabolic capabilities and to have genomic resources. In this context, the main objective of this research was to determine the genome features and the biochemical profile of *Synechococcus* sp. UCP002. The cyanobacterium was isolated from the Peruvian Amazon Basin region and cultured in BG-11 medium. Growth parameters, genome features, and the biochemical profile of the cyanobacterium were determined using standardized methods. *Synechococcus* sp. UCP002 had a specific growth rate of 0.086 ± 0.008 μ and a doubling time of 8.08 ± 0.78 h. The complete genome of *Synechococcus* sp. UCP002 had a size of ~3.53 Mb with a high coverage (~200x), and its quality parameters were acceptable (completeness = 99.29%, complete and single-copy genes = 97.5%, and contamination = 0.35%). Additionally, the cyanobacterium had six plasmids ranging from 24 to 200 kbp. The annotated genome revealed ~3,422 genes, ~3,274 protein-coding genes (with ~41.31% hypothetical protein-coding genes), two CRISPR Cas systems, and 61 non-coding RNAs. Both the genome and plasmids had the genes for prokaryotic defense systems. Additionally, the genome had genes coding the transcription factors of the metalloregulator ArsR/SmtB family, involved in sensing heavy metal pollution. The biochemical profile showed primary nutrients, essential amino acids, some essential fatty acids, pigments (e.g.,...
all-trans-β-carotene, chlorophyll a, and phycocyanin), and phenolic compounds. In conclusion, *Synechococcus* sp. UCP002 shows biotechnological potential to produce human and animal nutrients and raw materials for biofuels and could be a new source of genes for synthetic biological applications.

**KEYWORDS**
biotechnological exploitation, cyanobacteria, genome analysis, microbial biodiversity, nutraceuticals, biochemical analysis

### 1 Introduction

Cyanobacteria are an ancient lineage and polyphyletic group of prokaryotes exhibiting oxygenic photosynthesis (Kauff and Büdel, 2011). These microorganisms are the most important primary producers on Earth, inhabiting and playing key roles in a great diversity of aquatic and terrestrial ecosystems exposed to light (Dvořák et al., 2015). Estimates of the current cyanobacterial biodiversity range from 5,301 (AlgaeBase: Listing the World’s Algae, 2022) to 8,000 species (Guiry, 2012).

Due to their extraordinary biodiversity and metabolic diversity, cyanobacteria are potentially valuable for humans in several ways. For example, microalgae could be useful in aquaculture, producing foods, feeds, biofuels, fertilizers, nutraceuticals, secondary metabolites, pigments, and a myriad of bioactive biochemicals (Singh et al., 2005; Udayan et al., 2017; Cobos et al., 2020). In addition, cyanobacteria are isolable and cultivable in laboratory conditions (Ernst, 1991; Badr et al., 2018; Prihantini, 2020; Thilak et al., 2020; Jasser et al., 2022; Singh and Kumar, 2022). Moreover, cyanobacteria have a relatively short doubling time and can produce high biomass volumes (Yu et al., 2015; Jaiswal et al., 2018; Mastropetros et al., 2022; Prabha et al., 2022). These microorganisms also have small genomes readily deciphered at structural and functional levels (Jaiswal et al., 2018; Lin et al., 2019; Kling et al., 2022; Pierpont et al., 2022). Finally, several cyanobacteria are genetically transformable. The genetic transformability of these microorganisms can broaden their spectrum of biotechnological uses (Jaiswal et al., 2018; Jaiswal et al., 2020; Purdy et al., 2022; Santos-Merino et al., 2022; Tan et al., 2022).

Despite all these advantages of cyanobacteria as biotechnological platforms, only a few species have been exploited at commercial levels (Lem and Glick, 1985; Grewe and Pulz, 2012; Yu et al., 2015, 2). Consequently, it is necessary to constantly make bioprospection efforts and basic studies at the biochemical and molecular levels. These approaches will help us discover cyanobacterial strains with desirable phenotypic and genotypic traits. These desirable traits include the fastest growth, highest productivity, metabolic diversity, and genetic transformability (Thajuddin and Subramanian, 2005; Al-Haj et al., 2016; Selão, 2022). In this context, the present study shows the genome features and the biochemical profile of *Synechococcus* sp. UCP002, a cyanobacterium isolated from the Peruvian Amazon Basin region.

### 2 Materials and methods

#### 2.1 Sample collection

Water samples were collected horizontally along the water surface of the Amazon River using a 20-μm plankton net (Continental TEM, Lima, Peru) using a boat towing method. The plankton net was held horizontally at 20-cm depth and dragged ~100 m from the geographic coordinates 03°41′0.6″ S, 73°14′8.9″ W to 03°40′57″ S, 73°14′08″ W. Sterile, screw-cap, wide-mouth 500-ml glass bottles were used to collect and transport the water samples at ~8°C in dark conditions.

#### 2.2 Isolation, culture, growth profile, and harvest of the cyanobacterium

A total of 50 milliliters of the filtered water sample was homogenized with 50 ml of BG-11 medium (Allen, 1968). The cyanobacterial cells were cultured for 4 weeks in a controlled culture room at 25±0.06°C with 12:12-h light–dark cycles using 265 ± 10 μE m⁻² s⁻¹ intensity of a 50-W LED-based white light source (Wellmax, Samsung) with continuous bubbling of air and shaking the cultures at 180 rpm. After the initial cultivation of the mixed cultures, unicellular cyanobacteria were subjected to isolation by the cell washing method (Richmond and Hu, 2013) and by repeat subculturing and plating on a solid BG-11 agar medium.

After isolation, cyanobacterial cells were inoculated and cultured in 50 ml of the BG-11 medium. According to the growth of the cyanobacteria, the culture volume was increased gradually to obtain 1 L of culture.

To determine the cyanobacterium growth profile, an aliquot of the culture was taken and subcultured in triplicate at a final volume of 200 ml in Erlenmeyer flasks (250 ml) in a controlled culture room at 38.21 ± 0.76°C in 12:12-h light–dark cycles using 500 μE m⁻² s⁻¹ intensity of a 50-W LED-based white light source (Wellmax, Samsung), with continuous bubbling of air and shaking the cultures at 180 rpm. The initial absorbance at 730 nm (A₇₃₀ = 0.104 ± 0.05) was determined by
spectrophotometric analysis using a NanoDrop 2000 spectrophotometer (Thermo Fisher Scientific, United States). The growth of the cyanobacterium was monitored every day for 12 days by recording its absorbance at 730 nm and the dry weight of the cyanobacterial biomass. To determine both measures, aliquots of 2.1 ml were obtained from the culture every day at the sixth hour after the start of the illumination phase of the photoperiod. To determine the dry weight of the cyanobacterial biomass, 2 ml of the culture was harvested by centrifugation at 10,000 × g for 5 min at 4°C, and the pellets were washed two times with 1 ml of a physiological saline solution. Finally, the samples were dried in an oven at 70°C for 24 h, and the dry weight was measured gravimetrically using an analytical balance Kern ABJ 220-4NM (Kern and Sohn GmbH, Balingen, Germany).

To determine the cyanobacterium-specific growth rate (μ) and the doubling time (t_d, generation time in hours), an aliquot of the culture was subcultured in triplicate at a final volume of 200 ml in Erlenmeyer flasks (250 ml) in a controlled culture room at 38.21 ± 0.76 °C in constant illumination using 500 μE m⁻² s⁻¹ intensity of a 50-W LED-based white light source (Wellmax, Samsung), with continuous bubbling of air and shaking the cultures at 180 rpm. The initial absorbance at 730 nm was 0.085 ± 0.007. The growth of the cyanobacterium was monitored recording absorbances at 730 nm every 2 hours for 14 h. Based on the absorbance data on this period of time, the μ and t_d growth parameters were computed using the following equations:

\[
\text{Specific growth rate } (\mu) = \frac{\ln(A_f) - \ln(A_i)}{t_f - t_i},
\]

\[
\text{Doubling time } (t_d) = \frac{\ln(2)}{\mu},
\]

where A is the absorbance at 730 nm (A_730) at the final (f) or initial (i) time (t).

For the biochemical analysis, the cyanobacterial cells were harvested during the exponential growth phase of cultures at the sixth hour after the start of the illumination period. The culture was transferred to 50-ml conical-bottom centrifuge tubes and centrifuged at 2,000xg for 15 min at 4°C to harvest the cyanobacterial cells. The obtained cyanobacterial biomass was rinsed three times with 40 ml of sterilized ultrapure water, centrifuged again in the aforementioned conditions, and the supernatants were discarded.

### 2.3 Morphological and molecular identification of the isolated cyanobacterium

The isolated cyanobacterium was preliminarily identified using standard microscopic morphological characteristics. Also, the autofluorescence emitted was recorded using a Carl Zeiss fluorescence microscope. Microphotographs were obtained using a digital camera AxioLab.A1 AxioCam ERC real-time 5 s. Images were obtained at a magnification of ×400 with visible light and epifluorescence (excitation 510–560, emission 590). The average cell size (length and width) of the isolated cyanobacterium was estimated from 100 cells by ZEN 2012 × 32 blue software (Carl Zeiss, Jena, Germany).

For molecular identification, a phylogenomic analysis was conducted with 31 conserved proteins (Wu and Eisen, 2008). These conserved proteins were retrieved from complete genomes of *Synechococcus* sp. UCP002 and 43 cyanobacteria species. Furthermore, these proteins were concatenated, aligned, and trimmed using the tools of Geneious Prime™ 2022.2.2 (Kearse et al., 2012). Finally, a maximum likelihood tree with 100 bootstrap replicates was inferred using MEGA 11 (Tamura et al., 2021). The Le–Gascuel model (Le and Gascuel, 2008) of amino acid substitution was selected based on the likelihood test. A discrete gamma distribution was used to model evolutionary rate differences among sites (four categories (+G, parameter = 0.7090)) and a proportion of invariable sites.

### 2.4 Genomic DNA purification, library preparation, and shotgun sequencing

Genomic DNA was extracted from 200 mg of cyanobacterial biomass using a modified CTAB method (Cobos et al., 2017) and purified using the DNasey™ PowerSoil Pro Kit (QIAGEN, Germany), following the manufacturer’s instructions. The quality and quantity of the purified genomic DNA were determined by spectrophotometric analysis using a NanoDrop 2000 spectrophotometer (Thermo Fisher Scientific, United States). DNA integrity and purity were evaluated by electrophoretic analysis on agarose gels (Sambrook and Russell, 2006). DNA quantity was determined with the Qubit™ dsDNA BR Assay Kit using a Qubit™ 4 Fluorometer (Thermo Fisher Scientific, United States).

Libraries were prepared using the Nextera XT DNA Library Preparation Kit (Illumina, United States), following the manufacturer’s instructions. Purified DNA was fragmented and tagged using a tagmentation process. Index adapters were ligated to the tagmented DNA using a limited-cycle PCR program. The libraries were cleaned up by 0.8x Agencourt™ AMPure XP bead purification (Beckman Coulter, United States). The sizes of the libraries were determined using an Agilent High Sensitivity DNA Kit by Agilent 2100 Bioanalyzer microfluidic electrophoresis (Agilent Technologies, United States). Finally, the libraries were quantified using the Qubit™ dsDNA HS Assay Kit (Thermo Fisher Scientific) and paired-end sequenced with the Illumina NexSeq 550 platform.
2.5 Bioinformatic analysis

Raw Illumina paired-end reads were uploaded as FASTQ files and analyzed using the Galaxy (Jalili et al., 2020) and KBase (Arkin et al., 2018) platforms. From raw sequences, the high-quality reads were obtained with Trimmomatic v0.38.1 (Bolger et al., 2014), and read qualities were evaluated with FastQC v0.11.9 (Andrews, 2010).

To verify that the cyanobacterial strain is axenic or unaxenic, the clean reads were taxonomically assigned with GOTTCHA2 software v2.1.7 (Freitas et al., 2015). Because the results generated by GOTTCHA2 software show associated bacteria, the de novo assembly of the complete genome of Synechococcus sp. UCP002 was conducted, following the bioinformatic approaches described as follows.

Short sequences were subjected to the first round of the de novo assembly using the assemblers IDBA-UD v1.1.3 (Peng et al., 2012), MEGAHIT v1.2.9 (Li et al., 2015), metaSPAdes v3.15.3 (Nurk et al., 2017), SPAdes v3.15.3 (Bankevich et al., 2012), and Velvet v1.2.10 (Zerbino, 2010). Qualities and assembly parameters were assessed with QUAST v4.4 (Gurevich et al., 2013). Next, the second round of the de novo assembly was conducted using the totality of contigs obtained with the five assemblers. For this process, the contigs were elongated and assembled (scaffolding) using the mapper and the de novo assembler tools of Geneious Prime® 2022.2.2 (Kearse et al., 2012). Furthermore, the third round of the assembly was conducted using the MaSuRCA genome assembler v3.2.9 (Zimin et al., 2013) using a combination of the generated contigs and scaffolds and the high-quality short reads. In addition, to reconstruct the draft genome, contigs and scaffolds were binned using CONCOCT v1.1 (Alneberg et al., 2014), MaxBin2 v2.2.4 (Wu et al., 2016), and MetaBAT2 v1.7 (Kang et al., 2019). Binned contigs and scaffolds were optimized by dereplication, aggregation, and scoring approaches using the DAS Tool v1.1.2 (Sieber et al., 2018). Taxonomic assignments of the optimized bins were based on the Genome Taxonomy Database (GTDB; https://gtdb.ecogenomic.org/) (Parks et al., 2022) using GTDB-Tk v1.7.0 (Chauvel et al., 2020). The bin containing contigs and scaffolds derived from cyanobacteria (Synechococcus sp.) was extracted as an assembly using the BinnedContigs tool v1.0.2. Furthermore, the complete genome was obtained by re-assembling the contigs and scaffolds and the clean paired-end reads by Unicycler v0.4.8.0 software (Wick et al., 2017). Finally, prior to downstream annotation analysis, coverage, quality, contamination, and completeness of the genome were evaluated using the Geneious mapper (Kearse et al., 2012), CheckM v1.0.18 (Parks et al., 2015), and BUSCO v5.3.2 (Simão et al., 2015), respectively. Additionally, the prediction of plasmid sequences in contigs and scaffolds was conducted by PlasFlow v1.0 software (Krawczyk et al., 2018), and its assembly was completed by NOVOPlasty v4.3.1 software (Dierckxsens et al., 2017).

The circular genome map of Synechococcus sp. UCP002 was aligned with its closest genetic neighbors using the Proksee server (https://proksee.ca/). Also, the circular maps of the plasmids were generated using the same online server.

The assembled genome was functionally annotated using the following tools: Bakta v1.5.0 (Schwengers et al., 2021), dFast v1.6.0 (https://dfast.ddbj.nig.ac.jp/) (Tanizawa et al., 2018), DRAM v0.1.0 (Shaffer et al., 2020), KAAS (https://www.genome.jp/kaas/) (Moriya et al., 2007), Prokka v1.14.5 (Seemann, 2014), and RASTTk v1.073 (Brettin et al., 2015). Additionally, the CRISPR-Cas elements in the genome and plasmids were identified by the online software application CRISPRCasFinder (Covin et al., 2018).

The genes coding the enzymes of the phenylpropanoid/flavonoid biosynthetic pathway were found by conducting a local BLAST search (Camacho et al., 2009) according to Del Mondo et al. (2022). Sequences of 29 core enzymes acting in the phenylpropanoid/flavonoid biosynthetic pathway of plants and cyanobacteria from KEGG were used as queries to detect ortholog sequences in the complete genome of Synechococcus sp. UCP002. The sequences with the best hit matches for each core enzyme were retained. Furthermore, the protein sequences were used to make a BLAST search against the UniProt TrEMBL protein database (UniProt Consortium, 2021).

The protein sequences coded by the genes smtB of Synechococcus sp. UCP002 and some cyanobacterial species (Synechococcus sp. PCC 6312 (WP_015123347), Calothrix sp. PCC 7507 (WP_01512714), Leptolyngbya sp. PCC 6406 (WP_008314625), Oscillatoria nigro-viridis (WP_01574920), Nostoc sp. PCC 7107 (WP_015114276), and Anabaena sp. PCC 7108 (WP_016952607)) were aligned using the Alignment tool of Geneious Prime® 2022.2.2.

The prediction of the three-dimensional structure of the proteins SMTB1, SMTB2, and SMTA involved in the metal-responsive cyanobacterial expression system of Synechococcus sp. UCP002 was realized using the SWISS-MODEL server (https://swissmodel.expasy.org/). The three-dimensional models for the two proteins of the metal-sensing transcriptional repressors (SMTB) were based on the cyanobacterial metallothionein repressor from Synechococcus elongatus PCC 7942 (PDB accession: ISMT). The model corresponding to metallothionein was based on the cyanobacterial metallothionein SMTA from Synechococcus elongatus PCC 7942 (PDB accession: IJD).

2.6 Biochemical analysis of the cyanobacterial biomass

For proximate composition analysis, the cyanobacterial biomass was dried in an oven at 70°C. The dried biomass was measured gravimetrically using an analytical balance Kern ABJ 220-4NM (Kern and Sohn GmbH, Balingen, Germany). Total
lips were extracted following the Bligh and Dyer method (Bligh and Dyer, 1959) and quantified gravimetrically using a semi-micro analytical balance (Sartorius, MSU225S-000-DU, Foster City, CA, United States). Total carbohydrates were determined using a colorimetric method (DuBois et al., 1956). The protein content was measured following the Hartree approach (Hartree, 1972). The ash content was determined by thermogravimetry (AOAC, 1990) using a Thermolyne™ F6010 muffle furnace (Thermo Fisher Scientific, Waltham, MA, United States) set at 550°C for 16 h.

For pigment analysis (all-trans-β-carotene, lutein, and chlorophyll a), 40 mg of the freeze-dried cyanobacterial biomass was homogenized with 5 ml of acetone 100%. The aceton extract was filtered onto 0.45-μm PTFE membrane filters to remove cells and cell debris. Next, 20 μl of pigment solutions were resolved using a Hitachi Elite LaChrom HPLC System (Hitachi High Technologies, San Jose, CA, United States) equipped with an L-2200 autosampler, L-2130 HTA pump, L-2455 diode array detector, and a 150 × 4.6 mm x 5 μm MilliporeSigma™ LiChroCART™ LiChrosorb™ RP-8 C8 Reversed Phase HPLC Column (Merck, Darmstadt, Germany). The HPLC system was programmed to run under the following conditions: column temperature: 25°C, flow rate: 1 ml/min, and absorbance monitoring at 450 nm. A ternary mobile phase consisted of (A) 100% methanol, (B) methanol: ammonium acetate 0.5 N (80:20), and tetrahydrofuran. The following gradient elution was employed: 0 min: (0% A, 100% B, and 2% C), 26 min (98% A, 0% B, and 2% C), 34 min (0% B, and 0% C), 5 min (98% A, 0% B, and 2% C), 42.2 min (80% A, 0% B, and 2% C), and 30 min of column equilibration (100% A, 0% B, and 0% C). EZChrom Elite software v3.2.1 (Agilent Technologies, Santa Clara, CA, United States) was used for data acquisition and analysis, compared with the chromatographic profiles of authentic standards of all-trans-β-carotene, lutein, and chlorophyll a (Sigma-Aldrich, Saint Louis, MO, United States).

For the determination of the total content of phycocyanin in vivo (c-phycocyanin [CPC] + allophycocyanin [APC]), the absorbances at 620 and 652 nm of an aliquot of the cyanobacterial culture in the logarithmic growth phase were recorded using a NanoDrop 2000 spectrophotometer (Thermo Fisher Scientific, United States). Finally, the CPC and APC contents were determined with the following equations (Bennett and Bogorad, 1973; Chaiklahan et al., 2012):

\[
\text{CPC (mg mL}^{-1}) = \frac{A_{620} - (0.474 \times A_{403})}{5.34},
\]

\[
\text{APC (mg mL}^{-1}) = \frac{A_{652} - (0.208 \times A_{620})}{5.09}.
\]

For total phenolic content (TPC) analysis, first, a hydromethanolic extract was obtained from 100 mg of the cyanobacterial dry biomass using an approach previously described (Cobos et al., 2020). Furthermore, the total phenolic content was estimated by the Folin–Ciocalteu method (Velioglu et al., 1998) based on a standard curve from 10 to 100 μM of gallic acid (3,4,5-trihydroxy benzoic acid) (Sigma-Aldrich, Germany). Results of the total phenolic content were expressed as gallic acid equivalents (mg GAE. g⁻¹ of cyanobacterial biomass dry weight [cbdw]).

For fatty acid analysis, first, fatty acid methyl esters (FAMEs) were obtained following an acid-catalyzed methanolysis/methylation approach (Ichihara and Fukubayashi, 2010); furthermore, FAMEs were resolved by using a gas chromatographic method (Cobos et al., 2020). FAMEs were identified by comparing the retention time of the peaks with a known standard mixture (Nu-Chek Prep, Elysian, MN, United States). Also, each sample was mixed with tricosanoic acid methyl ester (Sigma-Aldrich, Saint Louis, MO, United States) as the internal standard. Finally, generated chromatograms were analyzed with Galaxie™ Chromatography Data System software v1.9.3.2 (Agilent Technologies, Santa Clara, CA, United States). For amino acid analysis, total proteins were subjected to acid hydrolysis (Hirs et al., 1954); furthermore, amino acids obtained by hydrolysis and amino acid standards (Sigma-Aldrich, Saint Louis, MO, United States) were derivatized with 6-aminoquinolyl-N-hydroxy succinimidyl carbamate, following instructions of the AccQ-Fluor Reagent Kit (Waters Corporation, Milford, MA, United States). Derivatized amino acids were identified and quantified using an HPLC method (Cohen and Michaud, 1993). All the described biochemical analyses were carried out in triplicate, and data are expressed as the mean ± SD.

3 Results and discussion

3.1 Isolation, growth profile, and identification

In this work, we report the discovery of the cyanobacterial strain Synechococcus sp. UCP002 from the Peruvian Amazon Basin. The unicellular cyanobacterium strain showed a typical growth profile with the lag, logarithmic, and stationary phases in an interval time of 12 days (Figure 1, Supplementary Figure S1). The average values for the specific growth rate and the cell doubling time in constant illumination for 14 h were estimated at 0.086 ± 0.008 μ and 8.08 ± 0.78 h, respectively (Figure 1). It is difficult to compare this cell doubling time value with reports for other cyanobacterial strains due to differences in culture conditions (e.g., photoperiod, light intensity, temperature, and use or number of photobioreactors). Some doubling time values are 4.9 h for Synechococcus elongatus PCC 7942 (Ungerer et al., 2018a) and 11.8 h for Synechococcus sp. AMCI49 (Mori et al., 1996; Kondo
et al., 1997). Also, some studies report the existence of fast-growing strains of *S. elongatus*. For example, the strains PCC 11801 and PCC 11802 isolated from Powai Lake (India) have doubling times of 2.3 h (Jaiswal et al., 2018) and 2.8 h (Jaiswal et al., 2020), respectively. Similarly, *Synechococcus* sp. PCC 11901 isolated from the Johor Strait (Singapore) has a doubling time of ≈2.0 h (Włodarczyk et al., 2020). Finally, *S. elongatus* UTEX 2973 is up-to-date the fastest-growing cyanobacterium with a doubling time of 1.5 h (Yu et al., 2015; Ungerer et al., 2018b).

To corroborate the cyanobacterial nature of the strain, first, the whole-cell absorbance spectrum was recorded in the UV-visible range (from 280 to 800 nm). The whole-cell absorbance showed a peak at 635 nm, which corresponds to C-phycocyanin (Abalde et al., 1998; Gupta and Sainis, 2010; Sonani et al., 2017). Additionally, the whole-cell absorbance showed peaks at 440 and 680 nm. This absorbance pattern is in agreement with the typical absorbance of chlorophyll a. Also, the whole-cell absorbance showed a less defined peak at 483 nm. This peak corresponds to carotenoids (Figure 1). Microscopically, the cyanobacterium is unicellular with rod-shaped morphology (Figure 1). The cyanobacterium had an average length of 3.61 ± 0.54 μm (from 2.49 to 4.34 μm) and a width of 1.42 ± 0.31 μm (from 1.06 to 2.10 μm). These cellular dimensions are similar to those of *Synechococcus* sp. 6,312 (2.7 × 1.3 μm) (Allen and Stanier, 1968) and some strains of *Synechococcus elongatus* (Jaiswal et al., 2018; Jaiswal et al., 2020). Together, these typical morphological characteristics corroborate that the cyanobacterium belongs to the Synechococcaceae family.

Finally, based on the phylogenomic analysis of 31 conserved proteins (Wu and Eisen, 2008), the isolated strain forms a clade with cyanobacteria of different genera. These cyanobacterial genera include *Acaryochloris*, *Synechococcus*, *Thermotrichus*, and *Termosynechococcus* (Figure 2). But *Synechococcus* sp. UCP002 showed the highest genetic similitude with *Synechococcus* sp. PCC 6312. *Synechococcus* sp. PCC 6312 is a fresh-water cyanobacterium isolated from California (United States) in 1963 (https://www.ncbi.nlm.nih.gov/biosample/SAMN02261337), and its complete genome was sequenced by the CyanoGEBA Sequencing Project (https://www.ncbi.nlm.nih.gov/bioproject/158717). *Synechococcus* sp. PCC 6312 shows similitude in shape (rod-shaped) and size (Allen and Stanier, 1968) with *Synechococcus* sp. UCP002. In addition, this cyanobacterium strain can intracellularly
biomineralize amorphous calcium carbonate. These calcium carbonate inclusions are located mostly in the cellular poles of the cyanobacterium (Benzerara et al., 2014; De Wever et al., 2019).

3.2 Genome analysis

3.3.1 Genome assembly

A total of 4,838,556 reads were de novo assembled to generate the complete genome of *Synechococcus* sp. UCP002. The complete genome had a size of ~3.53 Mb and a GC content of 47.87% (Figure 3). The GC content and the genome size recorded fit in the range of values reported for cyanobacteria with complete genomes of the genus *Synechococcus* (Supplementary Table S1) (Sugita et al., 2007; Jaiswal et al., 2018; Jaiswal et al., 2020, 11801; Kling et al., 2022; Pierpont et al., 2022, 1). Also, we assembled and annotated six plasmids with sizes ranging from 24.44 to 200.03 kbp (Supplementary Figure S2).

According to the analysis, the complete genome had a high coverage and was characterized by its high quality. The coverage was ~200x, and the BUSCO results indicate that more than 97% of the 40 core genes were complete and single-copy genes, and only a very low fraction was missing. In addition, the CheckM results showed that the genome had completeness of more than 99% and a low contamination level (0.35%).

FIGURE 2
Maximum likelihood phylogenomic tree with bootstrap support inferred based on 31 conserved proteins.
3.3.2 Functional annotation of the genome

The complete genome of *Synechococcus* sp. UCP002 contained 3,324–3,640 predicted genes (Table 1). Of the total genes, from 3,280 to 3,596 were protein-coding genes with an average protein length of 305 amino acid residues. Of these protein-coding genes, those from 1,920 to 2,407 were associated with known functions. Of the protein-coding genes with known functions, those from 756 to 948 were enzyme-coding genes. Also, from the total protein-coding genes, those from 903 to 1,752 were hypothetical protein-coding genes. In addition, in the draft genome were recognized seven ncRNA, nine ncRNA regions, two genes coding rRNA (16 and 23S), 41 genes coding tRNAs, one gene coding tmRNA, and two CRISPR–Cas systems. Also, CRISPR–Cas systems were identified in four (pSUCP002.1, pSUCP002.2, pSUCP002.3, and pSUCP002.6) of the six plasmids of *Synechococcus* sp. UCP002 (Supplementary Table S2). CRISPR loci and Cas (CRISPR-associated) operons together (CRISPR–Cas system) constitute a heritable molecular adaptive immune system found in many bacterial and archaean species (Barrangou et al., 2007; Sorek et al., 2008). In cyanobacteria, the CRISPR–Cas systems are classified into class I (with types I and III) and class 2 with type V (Pattharaprachayakul et al., 2020).

This CRISPR–Cas system protects the prokaryotic cells from invading bacteriophages and conjugative plasmids (Sorek et al., 2008; Marrafiini and Sontheimer, 2010). Recent investigations have shown that the CRISPR–Cas system has a widespread distribution in the phylum Cyanobacteria. However, marine cyanobacteria of the *Synechococcus* and *Prochlorococcus* subclade do not have this interference system (Cai et al., 2013). This result is apparently paradoxical because these marine cyanobacteria live in an environment with abundant and diverse cyanophages (Suttle and Chan, 1994; Lu et al., 2001).

Actually, the CRISPR–Cas system has emerged as an effective and versatile tool for genetic modification in cyanobacteria of the genus *Synechococcus*. Wendt et al., using targeted genome editing and enrichment outgrowth, created a new strain of *Synechococcus* elongatus 2973-1T that was both naturally transformable and fast-growing (Wendt et al., 2022). Also, some researchers demonstrated that the CRISPR–Cas tool can be used for the

| TABLE 1 Results of the functional annotation of the complete genome of *Synechococcus* sp. UCP002 using Bakta, dFast, Prokka, and RASTtk tools. |
|---------------------------------|
| Bakta | dFast | Prokka | RASTtk |
|---------------------------------|
| **Total gene count**           | 3,371 | 3,324 | 3,354 | 3,640 |
| **CRISPR array**               | 2     | 2     | 2     | 2     |
| **Protein-coding genes**       | 3,310 | 3,280 | 3,310 | 3,596 |
| **Number of genes with EC number** | 948  | 943  | 756  | 942  |
| **RNA**                        | 61    | 44    | 44    | 44    |
| **ncRNA**                      | 7     | —     | —     | —     |
| **ncRNA regions**              | 9     | —     | —     | —     |
| **rRNA**                       | 3     | 2     | 2     | 2     |
| **tRNA**                       | 41    | 41    | 41    | 41    |
| **tmRNA**                      | 1     | 1     | 1     | 1     |
| **Proteins associated with known function** | 2,407 | 1,920 | 1,558 | 2,111 |
| **Hypothetical proteins**      | 903   | 1,360 | 1,752 | 1,566 |
| **Coding ratio (%)**           | 85.9  | 84.9  | 85.6  | 86.7  |
metabolic engineering of cyanobacteria of the genus *Synechococcus* (Gordon et al., 2016; Racharaks et al., 2021). Consequently, this modern genetic tool could be used to generate genetically improved strains of *Synechococcus* sp. UCP002.

In addition to the CRISPR–Cas system, the genome of *Synechococcus* sp. UCP002 and most of its plasmids contain other prokaryotic defense systems (Supplementary Table S2). These include the restriction and modification system (RM system), the toxin–antitoxin system (TA system), and the DNA phosphorothioate system (PT). The RM system includes genes coding restriction enzymes (types I, III, and IV) and DNA methyltransferases (e.g., dam, yhdj, and dcm). The TA system has genes coding several key proteins of the type II TA system (e.g., mazE, prfA, yefM, and higA-1) and their related factors (i.e., ftsZ, mreB, and gltX). Finally, the PT system has genes coding the sulfur modification proteins (e.g., iscS, dnbB, dniK, and dndD) and genes coding the DNA phosphorothioation-dependent restriction proteins such as dptF, dptG, and dptH. The PT system modifies the DNA backbone and constitutes a protective epigenetic system with multiple functions (i.e., antioxidant, restriction-modification, and virus resistance properties). The occurrence of dnd genes and gene clusters is common in the genomes of archaea and bacteria, including cyanobacteria of the orders Nostocales and Synechococcales (Jian et al., 2021). Together, these prokaryotic defense systems are an evolutive response system of bacteria and archaea against the great diversity of genetic parasites. Consequently, in these microorganisms, an important fraction of the genetic information participates in antiparasitic defense. These antiparasitic defense systems use several strategies, including innate immunity (RM system), adaptive immunity (CRISPR–Cas system), dormancy induction, or programmed cell death (TA system) (Makarova et al., 2013; Koonin et al., 2017).

Two genes potentially useful for the development of biosensors for heavy metal pollution were identified in the genome of *Synechococcus* sp. UCP002, which are denominated *smtB1* and *smtB2*. These genes code transcription factors of the metalloregulator ArsR/SmtB family. These transcription factors are negative regulators for the expression of the gene *smtA*, also identified in the genome, which codes cysteine-rich metallothionein that is able to sequester metal ions, such as cadmium, copper, and zinc (Cavet et al., 2003). To corroborate the identity of the proteins encoded by *smtB1* and *smtB2*, we conducted additional *in silico* analysis by predicting their tridimensional structure (Supplementary Figure S3) and by aligning their amino acid sequences (Figure 4). According to this analysis, the two proteins fit with known proteins involved in the functions described. Together, these genes constitute a metal-responsive expression system (Turner et al., 1996; Busenlehner et al., 2003; Ma et al., 2009). This system is commonly used to develop biosensors to detect environmental pollution with heavy metals (Hui et al., 2021), and consequently, it will be necessary to conduct additional studies to verify the reliability of the genes *smtB1* and *smtB2* of *Synechococcus* sp. UCP002 in engineering biosensors for the detection of heavy metal pollution in our region.

### 3.3 Biochemical profiling

According to the bioinformatic analysis, 54% and 46% of the sequencing reads correspond to *Synechococcus* sp. PCC 6312 and the associated bacteria (e.g., Proteobacteria and Dietzia), respectively (Supplementary Figure S4). Consequently, the biochemical profile represents the values of the cyanobacterium and its associated bacteria.

#### 3.3.1 Proximate composition, pigments, and total phenolic content

The proximate composition of the unaxenic cyanobacterium showed significant differences in its organic and inorganic contents. Between the organic biomolecules, the proteins were the most abundant (>57%). The second most abundant biomolecules were carbohydrates (>17%), followed by lipids (~16%). In contrast, the inorganic content, constituted by ashes and moisture, was very low (<5%) (Figure 5). A similar composition of these biomolecules was previously reported for some cyanobacterial strains of the *Arthrospira* genera and microalgae isolated from the Peruvian Amazon Basin region (Cobos et al., 2020). It is necessary to consider that the biochemical composition of the cyanobacterial biomass depends on biotic and abiotic factors. The abiotic factors include the growth phase, the associate microorganisms (e.g., yeasts, microalgae, and bacteria), the presence of cyanophages, the genetic constitution of the strain, and its level of laboratory domestication (Wilson et al., 1996; Ni and Zeng, 2016; Adomako et al., 2022; Grasso et al., 2022). Also, the abiotic factors include the physical conditions of the culture (e.g., light intensity, photoperiod, and temperature) and the chemical composition of the culture medium (e.g., nitrogen source, salinity, N:P ratio, and CO2 levels) (Pathania and Srivastava, 2021). Consequently, the recorded proximate composition for the unaxenic cyanobacterium is relative and could change significantly. In other words, by obtaining an axenic strain of the cyanobacterium and changing the physicochemical conditions of the culture, the production of these main biomolecules could be modified. Nevertheless, regardless of using an unaxenic or axenic strain, the raw cyanobacterial biomass could be used to produce protein-enriched foods, biofuels (e.g., biodiesel and bioethanol), and several useful bioproducts through a biorefinery approach (Sivaramakrishnan et al., 2022; Wang et al., 2022).

The unaxenic cyanobacterium showed some typical cyanobacterial pigments with significant differences in their
The genome of *Synechococcus* sp. UCP002 also harbors most of the genes coding the enzymes responsible for the biosynthesis of chlorophyll a (Supplementary Figure S6). However, the genes coding for the enzymes for the biosynthesis of lutein, using lycopene as a metabolic precursor, were lacking in the genome. The missing genes coding the enzymes for the biosynthesis of lutein are typical in cyanobacteria of the *Synechococcus* genera (Sarnaik et al., 2018). Consequently, lutein identified in the unaxenic cyanobacterial biomass could be a result of its biosynthesis by the associated bacteria.

Additionally, the cyanobacterium strain showed the ability to produce phenolic compounds. The total phenolic content (TPC) recorded was 2.36 ± 0.06 mg GAE g⁻¹ of cyanobacterial biomass dry weight. This TPC value is in the range of values reported for TPC in cyanobacteria. A similar low TPC is recorded in strains of the cyanobacterial genera *Arthrospira* (Cobos et al., 2020). Also, Li et al. reported that *Synechococcus* sp. FACHB 283 has a TPC of 10.56 ± 0.11 mg GAE g⁻¹ of freeze-dried biomass weight (Li et al., 2007). In other cyanobacterial genera, such as *Aulosira*, *Anabaena*, *Aphanizomenon*, *Calothrix*, *Oscillatoria*, and *Synechocystis*, the TPC fluctuates from 22.17 to 290.23 mg GAE g⁻¹ of fresh biomass weight (Singh et al., 2017; Senousy et al., 2020). Additionally, Del Mondo et al. recorded in 14 cyanobacterial genera (e.g., *Anabaena*, *Chroococcus*, *Fischerella*, *Plectonema*, and *Tolyphothrix*) a TPC value from 1.0 to 60.53 mg GAE g⁻¹ of cbdw (Del Mondo et al., 2021).

Among the most common phenolic compounds identified in cyanobacteria are caffeic acid, chlorogenic acid, ferulic acid, gallic acid, protocatechuic acid, trans-cinnamic acid, *p*-coumaric acid, and vanillic acids (Babaoglu Aydaş et al., 2013; Jerez-Martel et al., 2017; Singh et al., 2017; Del Mondo et al., 2021).
the genome of *Synechococcus* sp. UCP002 has all the genes coding the enzymes of the shikimate pathway (Supplementary Figure S7). The shikimate pathway is the main aromatic biosynthetic pathway (Bentley, 1990; Mir et al., 2015). This metabolic pathway converts the metabolites erythrose-4-P and phosphoenolpyruvate to chorismate. Chorismate is the last common precursor for the biosynthesis of the three aromatic amino acids phenylalanine, tyrosine, and tryptophan (Herrmann, 1995). Furthermore, phenylalanine is the common metabolic precursor to biosynthesize multiple phenolic compounds through the phenylpropanoid biosynthetic pathway (Ferrer et al., 2008; Vogt, 2010).

However, it is intriguing that although the cyanobacterium produces phenolic compounds, not all the genes coding the 29 core enzymes of the phenylpropanoid/flavonoid biosynthetic pathway were found in the genome of *Synechococcus* sp. UCP002 (Supplementary Table S3). A total of 22 genes coding core enzymes were found in the genome but showed low pairwise identity with its corresponding orthologs (range from 21.4 to 39.3%). However, seven genes coding the core enzymes were missing in the genome. These missing genes are those that code for phenylalanine ammonia-lyase (EC 4.3.1.24), tyrosine ammonia-lyase (EC 4.3.1.25), caffeic acid 3-O-methyltransferase (EC 2.1.1.68), and chalcone synthase (EC 2.3.1.74). This suggests that *Synechococcus* sp. UCP002 could use other unidentified enzymes that convert phenylalanine or tyrosine into the metabolic intermediaries cinnamic acid or p-coumaric acid, respectively. An analogous situation could be occurring with the other missing core enzymes of the phenylpropanoid/flavonoid biosynthetic pathway. Recently, it has been demonstrated that cyanobacteria display greater variability in this metabolic pathway, and several core enzymes (e.g., phenylalanine ammonia-lyase, chalcone synthase, and

---

### TABLE 2 Fatty acid composition of the unaxenic cyanobacterium *Synechococcus* sp. UCP002 isolated from the Peruvian Amazon Basin region (mg.g\(^{-1}\) of cbdw).

| Fatty acid            | Mean   | Standard deviation |
|-----------------------|--------|--------------------|
| Saturated fatty acid  |        |                    |
| C14:0 (myristic acid) | 0.06   | 0.00               |
| C16:0 (palmitic acid) | 27.63  | 0.26               |
| C18:0 (stearic acid)  | 0.42   | 0.01               |
| Mono-unsaturated fatty acid (MUFA) | | |
| C16:1 n-7 (palmitoleic acid) | 34.51  | 0.31               |
| C18:1 n-7 (vaccenic acid) | 2.48   | 0.06               |
| C18:1 n-9 (elaidic or oleic acid) | 3.14   | 0.06               |
| Polysaturated fatty acid (PUFA) | | |
| C18:2 n-6 (linoleic acid) | 0.29   | 0.02               |
| Unknowns              |        |                    |
| Unknown 1             | 5.56   | 0.02               |
| Unknown 2             | 0.05   | 0.08               |
| Unknown 3             | 0.28   | 0.01               |
| Unknown 4             | 1.73   | 0.01               |
| Unknown 5             | 0.18   | 0.00               |
| ΣSFA                  | 28.11  | (36.83)            |
| ΣMUFA                 | 40.13  | (52.57)            |
| ΣPUFA                 | 0.29   | (0.38)             |
| Total fatty acids     | 76.33  |                    |

Legend: The concentration of fatty acids is expressed in mg/g of total lipids obtained from the unaxenic cyanobacterium dry biomass, with the percentage (%) of the total fatty acids in parentheses, and each value represents the mean ± SD of three experiments. ΣSFA is the sum of the contents of saturated fatty acids, ΣMUFA is the sum of the contents of mono-unsaturated fatty acids, and ΣPUFA is the sum of the contents of polysaturated fatty acids.
chalcone synthase) are missing in some analyzed cyanobacterial genera (e.g., Fischerella, Mastigocoleus, and Cylindrospermum) (Del Mondo et al., 2022).

3.1.2 Fatty acid composition

The unaxenic cyanobacterium contained three groups of fatty acids, according to their level of saturation, namely, saturated fatty acids (SFAs), monounsaturated fatty acids (MUFA)s, and polyunsaturated fatty acids (PUFAs) (Table 2). Palmitic acid (C16:0), palmitoleic acid (C16:1 n-7), and linoleic acid (C18:2 n-6) were the predominant ones in each fatty acid group. Comparing between groups, the content of MUFA`s was the highest (∑MUFA >52%), followed by the content of SFA (∑SFA >36%), but the content of PUFAs was the lowest (∑PUFA <0.5%). Also, five fatty acids were unknown (10.22% of total fatty acids). The pattern of fatty acid composition characterized by a high content of C16:0 and C16:1 n-7 fatty acids and low content of PUFAs is common in several cyanobacterial genera. These genera include Anabaena and Nostoc (Caudales and Wells, 1992; Guedes et al., 2011), Cyanobacterium (Sarsekeyeva et al., 2014), Dermocarpa, Dermocarpella, Myxosarcina, Pleurocapsa, Xenococcus (Caudales et al., 2000), Aphanothece, Oscillatoria, Pletonema, Phormidium (Oren et al., 1985), and Synechococcus (Murata et al., 1992; Takeyama et al., 1997; Yu et al., 2000; Santos-Merino et al., 2022). Consequently, these cyanobacteria do not produce the high-demanded polyunsaturated fatty acids EPA and DHA. These PUFAs are essential human nutrients biosynthesized for several eukaryotic microalgae (Nilsson et al., 2020; Hachicha et al., 2022; Jakhwal et al., 2022; Toumi et al., 2022).

The inability of Synechococcus sp. UCP002 to biosynthesize long-chain PUFAs can be associated with its genetic makeup for de novo fatty acid biosynthesis. According to the functional annotation with KAAS and the metabolic pathways reconstructed, the genome of the cyanobacterium possesses the totality of genes coding the enzymes involved in de novo fatty acid biosynthesis (Supplementary Figure S8). The type of the fatty acid biosynthesis pathway identified is II FAS, which typically operates in prokaryotic microorganisms such as cyanobacteria (Mund et al., 2022). Recently, it has been demonstrated that the enzymes of the fatty acid biosynthesis pathway form a protein community called metabolon (Skalidis et al., 2022). This metabolon is generated as the final product of the SFAs palmitic acid (C16:0) and stearic acid (C18:0). Furthermore, both saturated fatty acids are chemically modified by the sequential action of desaturases and elongases. Consequently, depending on the gene pool contained in the genome that codes desaturases and elongases, the cyanobacterium will have the metabolic competence to biosynthesize or not a group of short-chain and/or long-chain PUFAs (Ratledge, 2004; Poole et al., 2020; Santos-Merino et al., 2022).

In this context and taking into account the modes of fatty acid desaturation, the cyanobacterium Synechococcus sp. UCP002 belongs to groups 1 and 2, such as was previously established for this cyanobacterium genus (Murata et al., 1992). According to this, the genes coding elongases were missing in the genome of the cyanobacterium. However, in the genome were identified two genes coding desaturases, the first coding one delta-9 desaturase and the second coding one delta-12 desaturase. Together, these results suggest that the hydrocarbon chain length and the pattern of desaturation (desaturations at the Δ^9 and Δ^12 positions) of the fatty acids biosynthesized by Synechococcus sp. UCP002 were determined by the content of genetic information in its genome.

3.1.3 Amino acid composition

The unaxenic cyanobacterium strain contained 20 amino acids commonly found in proteins. In agreement with its nutritional value for human nutrition, it had both essential amino acids (EAA) and non-essential amino acids (NEAA).

| Amino acid | Mean | Standard deviation |
|------------|------|--------------------|
| Valine     | 35.06| 0.37               |
| Threonine  | 33.04| 0.45               |
| Leucine    | 54.72| 0.95               |
| Isoleucine | 33.03| 0.40               |
| Methionine + cysteine | 10.44 | 0.69         |
| Lysine     | 28.18| 1.25               |
| Histidine  | 9.32 | 0.16               |
| Phenylalanine | 30.94 | 0.45         |
| Tyrosine   | 27.88| 0.31               |

Legend: The quantity of amino acids is expressed in mg.g^{-1} of cyanobacterial biomass dry weight, with percentage (%) of amino acid content in parentheses, and each value represents the mean ± SD of three experiments. ΣEAA is the sum of the essential amino acids, and 2NEAA is the sum of the non-essential amino acids.
The content of both classes of amino acids in the cyanobacterial biomass was 45.62 and 54.38%, respectively (Table 3). Leucine was the main EAA, but histidine showed the lowest value (<10 mg g\(^{-1}\) of cbdw). In the NEAA group, glutamic acid (Glx) was more abundant and proline the scarcest (~24 mg g\(^{-1}\) of cbdw). Consequently, the *Synechococcus* sp. UCP002 biomass could be a good source of essential amino acids for human and animal nutrition. Essential amino acids are a common component in proteins derived from the cyanobacterial and microalgol biomass (Kay, 1991; Hempel et al., 2012; Molino et al., 2018; Camacho et al., 2019; Cobos et al., 2020).

The capability to biosynthesize the 20 amino acids by *Synechococcus* sp. UCP002 was supported by its genetic information. First, the genome harbors the genes coding specific permeases for ammonium assimilation and the involved enzymes to incorporate the ammonium ions into carbon skeletons (i.e., glutamine synthetase and glutamate synthase) (Muro-Pastor and Florencio, 2003; Muro-Pastor et al., 2005). Second, the genome possesses the nirABCD-narB operon (Supplementary Figure S9) for nitrate and nitrite assimilation (Omata et al., 1993). This operon is actively transcribed by nitrate and nitrite when the formation of glutamine decreases (Kikuchi et al., 1996). In *Synechococcus*, the best-characterized transcriptional activator of the nirABCD-narB operon is codified by the gene ntcA (Vega-Palas et al., 1992; Luque et al., 1994). Finally, the genome has the complete set of genes coding the enzymes for the biosynthesis of all amino acids.

For example, the biosynthesis of the aromatic family of amino acids (i.e., phenylalanine, tyrosine, and tryptophan) begins with the condensation of d-erythrose 4-phosphate and phosphoenolpyruvate. This condensation reaction produces inorganic phosphate and 3-deoxy-d-arabino-heptulosonate-7-phosphate (DAHP). The biochemical reaction is catalyzed by the enzyme phospho-2-keto-3-deoxyheptonate aldolase, also called DAHP synthase (EC 2.5.1.54). This enzyme catalyzes the first committed step of the shikimate pathway, which produces chorismate by the consecutive catalytic activity of several enzymes (Riccardi et al., 1989) (Supplementary Figure S7). DHAP synthase is the regulatory enzyme, and its feedback inhibition is by metabolic intermediates (i.e., phenylpyruvate, prephenate, and chorismate) and by aromatic end products (i.e., phenylalanine, tyrosine, and tryptophan). This regulatory mechanism by feedback inhibition is common in several genera of cyanobacteria such as *Chlorogloepsis, Fischerella, Lyngbya, Synechococcus, Synechocystis, Oscillatoria*, and *Plectonema* (Hall et al., 1982). Chorismate produced in the shikimate pathway is a fundamental metabolic branch point to biosynthesize the three aromatic amino acids by the consecutive action of several enzymes (Riccardi et al., 1989).

### 3 Conclusion

The cyanobacterium *Synechococcus* sp. UCP002 shows the potential to be biotechnologically exploited for the following reasons: 1) owing to their genetic makeup, the cyanobacterium can biosynthesize biochemicals potentially useful for human and animal nutrition, 2) some of the main biomolecules produced by the cyanobacterium (i.e., lipids and carbohydrates) could be used as raw materials to produce biofuels, and 3) the genomic resources of the novel cyanobacterium strain could be used in the field of synthetic biology as a new source of known genes with genetic variations.

### 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

MC and JC conceived, designed, and supervised the project. RC and MG collected the water samples and isolated the cyanobacterium. SE, MD, AT, and LB designed and carried out the experiments. JM, PA, and RR performed the biochemical analysis. CC, HR, and JV purified the genomic DNA, prepared libraries, and sequenced the genome. JC carried out the bioinformatics analysis and wrote the manuscript. All authors have read and approved the final manuscript.

### Funding

This research was supported by the Peruvian funding agency Consejo Nacional de Ciencia, Tecnología e Innovación Tecnológica (CONCYTEC) through the Programa Nacional de Investigación Científica y Estudios Avanzados (PROCIENCIA), Funding Award Contract No. 018-2018-FONDECYT/BM-Improvement of Research Infrastructure (Scientific Equipment).

### Acknowledgments

The authors thank the authorities and administrative–technical staff of the Universidad Científica del Perú and the Universidad Nacional de la Amazonía Peruana (UNAP) for providing us with facilities, to accessing to the laboratories, and their permanent support during the realization of the present investigation.
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/fgene.2022.973324/full#supplementary-material
for the analysis of hydrolysate amino acids via high-performance liquid chromatography. Anal. Biochem. 211, 279–287. doi:10.1016/0003-2697(93)20216-B

Cobos et al. 10.3389/fgen.2022.973324

Dordrecht: Springer Netherlands), 707

Ecology of cyanobacteria II: Their diversity in space and time

assessment tool for genome assemblies.

C., et al. (2022). Biomolecules from microalgae and cyanobacteria: Applications and market survey. Appl. Sci. 12, 1924. doi:10.3390/app121924

Hall, G. C., Flick, M. B., Gherna, R. L., and Jensen, R. A. (1982). Biochemical diversity for biosynthesis of aromatic amino acids among the cyanobacteria. J. Bacteriol. 149, 65–78. doi:10.1128/jb.149.1.65-78.1982

Hartree, E. F. (1972). Determination of protein: A modification of the lowry method that gives a linear photometric response. Anal. Biochem. 68, 422–427. doi:10.1016/0003-2697(72)90049-2

Hempel, N., Petrík, I., and Behrendt, F. (2012). Biomass productivity and productivity of fatty acids and amino acids of microalgal strains as key characteristics of suitability for biodiesel production. J. Appl. Phycol. 24, 1407–1418. doi:10.1007/s10811-012-9795-3

Herrmann, K. M. (1995). The shikimate pathway as an entry to aromatic secondary metabolism. Plant Physiol. 107, 7–12. doi:10.1104/pp.107.7.176

Hirs, C. H., Stein, W. H., and Moore, S. (1954). The amino acid composition of ribonuclease. J. Biol. Chem. 211, 941–950. doi:10.1016/0021-9258(58)11181-2

Hui, C., Guo, Y., Liu, L., and Yi, J. (2021). Recent advances in bacterial biosensing and biomaterials for enhanced biocatalysis. FEMS Microbiol. Rev. 35, 203–226. doi:10.1093/femsre/fuaa111

Ichikara, K., and Fukuyabashi, Y. (2010). Preparation of fatty acid methyl esters for gas-liquid chromatography. J. Lipid Res. 51, 635–640. doi:10.1194/jlr.D010106

Jaiswal, D., Sengupta, A., Sengupta, S., Madhu, S., Pakrasi, H. B., and Wangigkar, P. P. (2020). A novel cyanobacterium Synechococcus elongatus PCC 11802 has distinct genetic and metabolic characteristics compared to its neighbor PCC 11801. Sci. Rep. 10, 191. doi:10.1038/s41598-019-57051-0

Jaiswal, D., Sengupta, A., Soloniti, S., Sengupta, S., Phadnavis, A. G., Pakrasi, H. B., et al. (2018). Genome features and biochemical characteristics of a robust, fast growing and naturally transformable cyanobacterium Synechococcus elongatus PCC 11801 isolated from India. Sci. Rep. 8, 16632. doi:10.1038/s41598-018-34872-4

Jahwalkh, P., Kumar Biswas, J., Tiwari, A., Kwon, E. E., and Bhatnagar, A. (2022). Genetic and non-genetic tailoring of microalgae for the enhanced production of eicosapentaenoic acid (EPA) and docosahexaenoic acid (DHA) – a review. Bioresour. Technol. 344, 126250. doi:10.1016/j.biortech.2021.126250

Jalili, V., Afagian, E., Gu, Q., Clements, D., Blankenberg, D., Gneec, J., et al. (2020). The Galaxy platform for accessible, reproducible and collaborative biological analyses: 2020 update. Nucleic Acids Res. 48, W395–W402. doi:10.1093/nar/gkaa343

Jasser, I., Panou, M., Khomutovskia, N., Sandzewicz, M., Panteris, E., Niyatbekov, T., et al. (2022). Cyanobacteria in hot pursuit: Characterization of cyanobacteria strains, including novel taxa, isolated from geothermal habitats from different ecosystems of the world. Mol. Phylogenet. Evol. 170, 107454. doi:10.1016/j.ympev.2022.107454

Jerez-Martel, I., Garcia-Piza, S., Rodrigo-Martel, G., Rico, M., Afonso-Oliveiras, C., and Gómez-Pinçheti, J. L. (2017). Phenolic profile and antioxidant activity of crude extracts from microalgae and cyanobacteria strains. J. Food Qual. 2017, 1–8. doi:10.1155/2017/2924508

Jia, H., Xu, G., Yi, H., Yao, W., Yang, X., Xiong, L., et al. (2021). The origin and impeded dissemination of the DNA phosphoribosylation system in prokaryotes. Nat. Commun. 12, 6382. doi:10.1038/s41467-021-26636-7

Kang, D. D., Li, F., Kirtom, E., Thomas, A., Egan, R., An, H., et al. (2019). MetaBAT 2: An adaptive binning algorithm for robust and efficient genome reconstruction from metagenome assemblies. PeerJ 7, e7359. doi:10.7717/peerj.7359

Kauf, F., and Bäud, B. (2011). “Phylogeny of cyanobacteria: An overview,” in Progress in botany 72 progress in botany. Editors U. E. Lüttge, W. Beyschlag, B. Bäud, and D. Francis (Berlin, Heidelberg: Springer), 209–224. doi:10.1007/978-3-642-13455-5_8

Keay, R. A. (1991). Microalgae as food and supplement. Crit. Rev. Food Sci. Nutr. 30, 553–573. doi:10.1080/10408399109527556

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/bts099

Kikuchi, H., Aichi, M., Suzuki, I., and Oamato, T. (1996). Positive regulation by nitrile of the nitrate assimilation operon in the cyanobacterium Synechococcus sp. strain PCC 7942 and Plectonema boryanum. J. Bacteriol. 178, 5822–5825. doi:10.1128/jb.178.11.5822-5825.1996

Kling, J. D., Webb, E. A., and Hutchins, D. A. (2022). Genome sequence of Synechococcus sp. strain LA31, isolated from a temperate estuary. Microb. Resour. Announc. 11, e007721–21. doi:10.1128/mra.00775-21

Kondo, T., Mori, T., Lebedeva, N. V., Aoki, S., Ishiura, M., and Golden, S. S. (1997). Circadian rhythms in rapidly dividing cyanobacteria. Science 275, 224–227. doi:10.1126/science.275.5297.224
Thajuddin, N., and Subramanian, G. (2005). Cyanobacterial biodiversity and potential applications in biotechnology. Curr. Sci. 89, 47–57.

Thalak, T. S., Madhusoodanan, P. V., Pradeep, N. S., and Prakashumar, R. (2020). Isolation and taxonomy of the blue-green algae (cyanobacteria), Nostoc and Anabaena in Kerala state, India. Acta Bot. Hung. 62, 163–174. doi:10.1556/034.62.2020.1-2.10

Touni, A., Politsea, N., Dvoruc, S., Mukhatemova, L., and Ilanyashen, S. (2022). Obtaining DHA-EPA oil concentrations from the biomass of microalgae chlorella sorokiniana. Resources 11, 20. doi:10.3390/resources1102020

Turner, J. S., Glands, P. D., Samson, A. C. R., and Robinson, N. J. (1996). Zn 2+ sensing by the cyanobacterial metallothionein repressor SmrB. Different motifs mediate metal-induced protein-DNA dissociation. Natl. Acad. Sci. 93, 7314–7317. doi:10.1093/nar/25.14.3174

Udayan, A., Arumugam, M., and Pandey, A. (2017). “Chapter 4 - nutraceuticals from algae and cyanobacteria.” In Algal green chemistry (Amsterdam: Elsevier), 65–89. doi:10.1016/B978-0-44-457849-0-00004-7

Ungerer, J. Lin, P.-C., Chen, H.-Y., and Pakrasi, H. B. (2018a). Adjustments to photosystem stoichiometry and electron transfer proteins are key to the remarkably fast growth of the cyanobacterium Synechococcus elongatus UTEX 2973. mBio 9, e03257–17. doi:10.1128/mBio.03257–17

Ungerer, J., Wendt, K. E., Hendry, J. I., Maranas, C. D., and Pakrasi, H. B. (2018b). Comparative genomics reveals the molecular determinants of rapid growth of the cyanobacterium Synechococcus elongatus UTEX 2973. Proc. Natl. Acad. Sci. U. S. A. 115, E1761–E1770. doi:10.1073/pnas.1419212115

Vega-Palasa, M. A., Flores, E., and Herrero, A. (1992). NtcA, a global nitrogen regulator from the cyanobacterium Synechococcus that belongs to the Crp family of bacterial regulators. Mol. Microbiol. 6, 1853–1859. doi:10.1111/j.1365-2958.1992. tb01357.x

Velagiou, Y. S., Mazza, G., Gao, L., and Oomah, B. D. (1998). Antioxidant activity and total phenolics in selected fruits, vegetables, and grain products. J. Agric. Food Chem. 46, 4113–4117. doi:10.1021/jf9801973

Vogt, T. (2010). Phenylpropanoid biosynthesis. Mol. Plant 3, 2–20. doi:10.1093/ mp/mps106

Wang, S., Mukhamber, Y., Esakkomuthu, S., and Abomohra, A.-E. F. (2022). Integrated microbial biorefinery – routes, energy, economic and environmental perspectives. J. Clean. Prod. 348, 131245. doi:10.1016/j.jclepro.2022.131245

Watanabe, M., and Ikeuchi, M. (2013). Phycobilisome: Architecture of a light-harvesting supercomplex. Photosyn. Res. 116, 265–276. doi:10.1007/s11120-013-9905-3

Wendt, K. E., Walker, P., Sengupta, A., Ungerer, J., and Pakrasi, H. B. (2022). Engineering natural competence into the fast-growing cyanobacterium Synechococcus elongatus strain UTEX 2973. Appl. Environ. Microbiol. 88, e018221–21. doi:10.1128/AEM.01822-21

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, W. H., Carr, N. G., and Mann, N. H. (1996). The effect of phosphate status on the kinetics of cyanophyte infection in the oceanic cyanobacterium Synechococcus sp. Algal Biol. 3, 215. doi:10.1016/0922-6071(96)00056-x

Wlodarczyk, A., Selio, T. T., Norling, B., and Nixon, P. J. (2020). Newly discovered Synechococcus sp. PCC 11901 is a robust cyanobacterial strain for high biomass production. Commun. Biol. 3, 215. doi:10.1038/s42003-020-0910-0

Wu, M., and Eisen, J. A. (2008). A simple, fast, and accurate method of phylogenetic inference. Genome Biol. 9, R151. doi:10.1186/gb-2008-9-10-r151

Xu, Y.-W., Simmons, B. A., and Singh, S. W. (2016). MaxBln 2.0: An automated binning algorithm to recover genomes from multiple metagenomic datasets. Bioinformatics 32, 605–607. doi:10.1093/bioinformatics/btw638

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 biosynthetic using light and CO2. Sci. Rep. 5, 8132. doi:10.1038/srep08132

Yu, R., Yamada, A., Watanabe, K., Yazawa, K., Takeyama, H., Matsunaga, T., et al. (2000). Production of eicosapentaenoic acid by a recombinant marine cyanobacterium Synechococcus sp. Lipids 35, 1061–1064. doi:10.1007/s11775-000-0619-6

Zerbino, D. R. (2010). Using the Velvet de novo Assembler for Short-Read Sequencing Technologies. Curr. Protoc. Bioinform. 31, 11.5. doi:10.1002/0471295053.bivt10531

Zimin, A. V., Marçais, G., Puiu, D., Roberts, M., Salberg, S. L., and Yorke, J. A. (2013). The MaSuLCG genome assembler. Bioinformatics 29, 2669–2677. doi:10.1093/bioinformatics/bt347