Comparative Genomic Analysis Revealed Distinct Molecular Components and Organization of CO₂-Concentrating Mechanism in Thermophilic Cyanobacteria

Jie Tang¹, Huizhen Zhou¹, Dan Yao¹, Sadaf Riaz², Dawei You², Anna Klepacz-Smółka³ and Maurycy Daroch²*¹

¹ School of Food and Bioengineering, Chengdu University, Chengdu, China, ² School of Environment and Energy, Peking University Shenzhen Graduate School, Shenzhen, China, ³ Department of Bioprocess Engineering, Faculty of Process and Environmental Engineering, Łódź University of Technology, Łódź, Poland

Cyanobacteria evolved an inorganic carbon-concentrating mechanism (CCM) to perform effective oxygenic photosynthesis and prevent photorespiratory carbon losses. This process facilitates the acclimation of cyanobacteria to various habitats, particularly in CO₂-limited environments. To date, there is limited information on the CCM of thermophilic cyanobacteria whose habitats limit the solubility of inorganic carbon. Here, genome-based approaches were used to identify the molecular components of CCM in 17 well-described thermophilic cyanobacteria. These cyanobacteria were from the genus Leptodesmis, Leptolyngbya, Leptothermofonisia, Thermoleptolyngbya, Thermotrichus, and Thermosynechococcus. All the strains belong to β-cyanobacteria based on their β-carboxysome shell proteins with 1B form of Rubisco. The diversity in the Cᵢ uptake systems and carboxysome composition of these thermophiles were analyzed based on their genomic information. For Cᵢ uptake systems, two CO₂ uptake systems (NDH-1³ and NDH-1⁴) and BicA for HCO₃⁻ transport were present in all the thermophilic cyanobacteria, while most strains did not have the Na⁺/HCO₃⁻ Sbt symporter and HCO₃⁻ transporter BCT1 were absent in four strains. As for carboxysome, the β-carboxysomal shell protein, ccmK2, was absent only in Thermoleptolyngbya strains, whereas ccmK3/K4 were absent in all Thermotrichus and Thermosynechococcus strains. Besides, all Thermotrichus and Thermosynechococcus strains lacked carboxysomal β-CA, ccaA, the carbonic anhydrase activity of which may be replaced by ccmM proteins as indicated by comparative domain analysis. The genomic distribution of CCM-related genes was different among the thermophiles, suggesting probably distinct expression regulation. Overall, the comparative genomic analysis revealed distinct molecular components and organization of CCM in...
INTRODUCTION

Thermophilic cyanobacteria are photoautotrophic prokaryotes with a cosmopolitan distribution in diverse thermal environments (Tang et al., 2018a,b; Alcorta et al., 2020). With the increase of studies on these unique microorganisms and their niches, their importance as primary photosynthetic producers of the geothermal ecosystems becomes apparent. The thermophilic cyanobacteria account for a large part of the thermal ecosystem biomass and productivity. From an applicative perspective, thermophilic cyanobacteria can be explored for high value-added products relevant to agricultural, pharmaceutical, nutraceutical, and industrial applications (Patel et al., 2019). Particularly, the potential contribution of thermophilic cyanobacteria to mitigating carbon dioxide becomes attractive to researchers in the context of global warming and greenhouse gas emissions. Thermophilic cyanobacteria may facilitate the collection of CO₂ right from locations where it was generated, e.g., fossil fuels power plants, thanks to their higher thermostability of photosynthetic apparatus and resistance to photosynthetic inhibitors such as NO₃ and SO₄ found in high levels in the flue gases (Liang et al., 2019; Chen and Wang, 2020).

The CO₂-concentrating mechanism (CCM) is a vital biological process for CO₂ utilization by cyanobacteria under conditions of CO₂ limitation. Cyanobacterial CCM enables adequate photosynthetic CO₂ fixation by elevating the CO₂ level near the active site of Rubisco, thus providing adaptation to various CO₂-limited niches (Clement et al., 2016; Raven et al., 2017). For most thermophilic cyanobacteria, the aquatic environments where they live suffer from low availability of inorganic carbon (Ci), primarily due to external factors, e.g., temperature, pH and gas exchange (Durall and Lindblad, 2015).

The dissolution of gaseous CO₂ in water starts from CO₂ (g) transferring to CO₂ (aq), i.e., CO₂ mass transfer from gas to liquid phase (Figure 1A). Further, the small fraction of dissolved CO₂ reacts in the liquid phase via two parallel paths depending on pH and subsequent availability of OH⁻. One leads to H⁺ and HCO₃⁻ formation with the unstable intermediate product – carbonic acid H₂CO₃⁺, while the second with hydroxide ions also forms bicarbonate ions which further can dissociate into H⁺ and CO₃⁻. When the pH of the aqueous solution is lower than 8, CO₂ hydration follows the first path, and due to the very low value of the dissociation constant of the carbonic acid, the formation of carbonate ions can be neglected (Kierzkowska-Pawlak et al., 2022). The direct data on CO₂ dissociation and different species of inorganic carbon present in hot springs have not been much discussed in the literature yet. However, the known equations of dissociation constants of CO₂ in water and saline solutions are sufficient approximations to understand what inorganic carbon species are available across different pH ranges of those hot springs. The most commonly used equations for dissociation constants of CO₂ in water (both fresh and saline) are applicable up to 40–50°C (Mook, 2000; Woosley, 2021). Based on those works, Bjerrum plot (Figure 1) of the carbonate system can be further extrapolated to 72°C and adjusted for salinity (Mook, 2000; Solomon, 2001; Tremaine et al., 2004). The solubility of CO₂ in water is a function of temperature and do lower extend salinity (Figure 1), the higher the temperature, the lower the carbon dioxide solubility, elevating the temperature from 30 to 60°C results in 40% lower availability of CO₂. Further decrease of that solubility occurs due to increased salinity or presence of other dissolved elements ubiquitous in hot springs. Meanwhile the distribution of carbon species across different pH is also affected by the temperature and to lesser extend salinity (Figure 1). Whilst bicarbonate (HCO₃⁻) remains a dominant dissolved inorganic specie between pH 6 and 10; its ratio with other carbon species is temperature sensitive. The equilibrium between carbonic acid, bicarbonate and carbonate species is acid-shifted with an increase of temperature. Therefore, to survive in their unique aquatic habitat, the thermophilic cyanobacteria utilize CCM to ensure that the Rubisco with low affinity for CO₂ is surrounded by high CO₂ levels and regularly function regardless of thermal stress.

Prior studies on the effect of temperature on Rubisco, including that of thermophilic origin revealed that the maximal rates of carboxylation reaction are typically higher than optimal growth temperatures of an organism (Galmés et al., 2015, 2016). This discrepancy in is partially due to the decreased affinity of Rubisco and lower specificity of carboxylation versus oxidation reaction at higher temperatures (S_C grazing) (Gubernator et al., 2008; Galmés et al., 2015, 2016). All in all thermophilic cyanobacteria must operate carbon concentration mechanisms to mitigate both challenges associated with lower carbon availability resulting from its poor solubility, and lower specificity of Rubisco toward CO₂ at elevated temperatures.

In general, the cyanobacterial CCM comprises two primary components: Ci uptake systems and carboxysomes. Cyanobacteria have been reported to have up to five different systems to actively acquire and transport Ci into the cells, namely two for uptake of CO₂ and three for transport of HCO₃⁻ (Pronina et al., 2017). The two CO₂ uptake systems that convert cytosolic CO₂ into HCO₃⁻ rely on plastoquinone oxidoreductase NADPH dehydrogenase respiratory complexes (Price, 2011). The complexes include NDH-1, a low-CO₂ inducible high-affinity CO₂ uptake system encoded by ndhD3, ndhF3 and cupA (chpY) genes, and NDH-14, a constitutive low-affinity CO₂.
uptake system encoded by *ndhD4*, *ndhF4*, and *cupB* (*chpX*) genes (Kaplan, 2017). The transport of HCO$_3^-$ takes place at the plasma membrane and is performed by three transporters, including *bicA* (a *sulP*-type sodium-dependent HCO$_3^-$ transporter), *sbtA* (a sodium-dependent HCO$_3^-$ symporter), and *BCT1* (an ATP-binding cassette ABC-type HCO$_3^-$ transporter) (Raven et al., 2017). Inside the cyanobacterial cell, the HCO$_3^-$ is further transported into the carboxysomes that comprise protein shells and two encapsulated enzymes, Rubisco and carbonic anhydrase (CA) (Kerfeld and Melnicki, 2016). CA catalyzes the conversion of HCO$_3^-$ into CO$_2$, which is a substrate for Rubisco. Rubisco consisting in cyanobacteria of eight small (*rbcS*) and eight large (*rbcL*) subunits, catalyzes CO$_2$ fixation reaction to generate 3-phosphoglycerate for subsequent rearrangement through the Calvin-Benson-Bassham cycle (Hill et al., 2020). There are two types of shell proteins encoded by *cso* operon and *ccmKLMNO* operon, termed α-carboxysomes and β-carboxysomes, respectively. Based on this criterion, the cyanobacterial species carrying form 1A of Rubisco within α-carboxysomes are classified as α-cyanobacteria, while the species containing form 1B of Rubisco within β-carboxysomes as β-cyanobacteria (Melnicki et al., 2021). The *ccmP* was another β-carboxysome component, forming a bilayered shell protein (Cai et al., 2013). In addition, various carboxysomal CAs have been reported, including β-CA (*ccaA*) and γ-CA (*ccmM*) in β-cyanobacteria and β-CA (*csoSCA*) in α-cyanobacteria (MacCready et al., 2020). Two types of non-carboxysomal CAs were also reported in β-cyanobacteria, namely *ecaB* (β-CA) and *ecaA* (α-CA) (Sun et al., 2019).

Although CCM of freshwater, marine and alkaliphilic cyanobacteria have been reported (Burnap et al., 2015; Kupriyanova and Samyлина, 2015; Klanchui et al., 2017), no comprehensive investigation on CCM of thermophilic cyanobacteria have been performed. Moreover, previous studies mainly focused on cyanobacterial CCM related to environments with distinct C$_3$ content, salinity and especially pH that strongly influenced the CCM due to the equilibrium of C$_3$ species (Mangan et al., 2016). Therefore, it is indispensable to survey the CCM of thermophilic strains in relation to their habitats that limit the solubility of inorganic carbon (Kamennaya et al., 2012). In recent years, next-generation sequencing (NGS) has been extensively employed to investigate the cyanobacterial genomes in thermal environments (Walter et al., 2017; Alcorta et al., 2020; Chen et al., 2021). This offers an opportunity to explore the thermophilic cyanobacterial CCM and the relationship with their niches at the genomic level. Furthermore, the studies on the molecular components of cyanobacterial CCM in relation to their specific habitats may improve photosynthetic CO$_2$ fixation and biomass production in cyanobacteria (Noreña-Caro and Benton, 2018) and crop plants (Hennacy and Jonikas, 2020) through genetic engineering.

In the present study, genome sequences of 17 thermophilic cyanobacterial strains were used to investigate the molecular basis of CCM by computational identification. The relationship between CCM components and adaptations of these thermophilic cyanobacteria were further discussed. The insights into the CCM components lay a solid foundation for future research regarding photosynthetic improvement and biomass yield of thermophilic cyanobacteria with biotechnological potentials.

**MATERIALS AND METHODS**

**Genome Collection of Thermophilic Cyanobacteria**

According to the genomic resources of the NCBI at the time of this study (2021/05/20), cyanobacteria with available genomes and closely related to thermophilic or hot-springing strains were first retrieved as a preliminary dataset. To ensure the accurate collection of thermophilic cyanobacteria, these strains were further filtered by conducting literature searches, contacting culture collections, leading authors of the manuscripts and submissions to retain thermophilic cyanobacteria with definite information about their thermophilic characteristics. Then, quality control for all genomes was performed based on the following criteria to reduce data redundancy and biased genome representation of cyanobacteria. First, the whole-genome average nucleotide identity (ANI) for each pair of genomes was calculated using the ANI calculator with default settings$^1$. Genomes with an ANI value greater than 99.9% were considered as redundant genomes, and then only one of

---

$^1$[http://enve-omics.ce.gatech.edu/ani/](http://enve-omics.ce.gatech.edu/ani/)
these genomes was randomly kept for the analysis. Second, the quality of the genomes was evaluated using CheckM (Parks et al., 2015) to ensure a more reliable genome dataset with near completeness (≥95%) and low contamination (<5%). These analyses finally generated a dataset comprising 16 thermophilic cyanobacteria. The genome, protein sequences and genomic annotations of the thermophilic cyanobacteria studied were downloaded from the database of NCBI. The genomes with no or incomplete annotations were annotated using the RAST annotation system (Aziz et al., 2008), provided in Supplementary Table 1.

Information regarding the 16 thermophilic strains was summarized in Table 1. Briefly, the 16 thermophiles were affiliated to four families of order Pseudanabaenales and Synechococcales, including Leptolyngbyaceae: *Leptodesmis sichuanensis* A121 (Tang et al., 2022a), *Leptolyngbya* sp. JSC-1 (Brown et al., 2010), *Leptothermofonsia sichuanensis* E412 (Tang et al., 2022b); Oculatellaceae: *Thermoleptolyngbya* sp. O-77 (Yoon et al., 2017), *T. sichuanensis* A183 (Tang et al., 2021); Synechococcaceae: *Synechococcus* sp. 60AY4M2, 63AY4M2, 65AY6A5, and 65AY6Li (Olsen et al., 2015), *Thermostichus* sp. JA-2-3B and JA-3-3Ab (Bhaya et al., 2007); Thermosynechococcaceae: *Thermosynechococcus* sp. CL-1 (Cheng et al., 2020) and TA-1 (Leu et al., 2013), *T. vestitus* BP-1 (Nakamura et al., 2002) and E542 (Liang et al., 2019), *T. vulcanus* NIES-2134 (Liang et al., 2018).

### Table 1: Ecological information and genome characteristics of thermophilic cyanobacteria studied.

| Species                        | Isolation source                        | Niche temperature (°C) | Niche pH | Genome completeness (%) | Genome contamination (%) | Accession number | References                      |
|-------------------------------|-----------------------------------------|------------------------|----------|-------------------------|--------------------------|------------------|--------------------------------|
| *Leptodesmis sichuanensis* A121 | Hot spring, Erdaoqiao, Sichuan, China   | 40.8°C                 | 6.32     | 99.53                   | 0.94                     | GCA_021379005    | Tang et al., 2022a            |
| Leptolyngbya sp. JSC-1         | La Duke Hot Springs, Montana, United States | 60°C                  | 6.85     | 99.53                   | 1.30                     | GCA_000733415    | Brown et al., 2010            |
| *Leptothermofonsia sichuanensis* E412 | Hot spring, Lotus Lake, Sichuan, China | 42.7°C                 | 8.61     | 99.29                   | 0.00                     | GCA_019891175    | Tang et al., 2022b            |
| *Thermoleptolyngbya* aichuanensis A183 | Hot spring, Erdaoqiao, Sichuan, China | 40.8°C                 | 6.32     | 98.94                   | 0.94                     | GCA_000733415    | Brown et al., 2020            |
| *Thermoleptolyngbya* sp. O-77  | Hot spring, Kumamoto, Japan              | 35–60°C                | NA       | 98.70                   | 1.53                     | GCA_001548395    | Yoon et al., 2017             |
| *Thermostichus* sp. 60AY4M2    | Mushroom Spring, Yellowstone National Park, United States | 60°C                  | alkaline | 100.00                  | 0.88                     | GCA_002760375    | Olsen et al., 2015            |
| *Thermostichus* sp. 63AY4M2    | Mushroom Spring, Yellowstone National Park, United States | 63°C                  | alkaline | 100.00                  | 0.00                     | GCA_002760475    | Olsen et al., 2015            |
| *Thermostichus* sp. 65AY6A5    | Mushroom Spring, Yellowstone National Park, United States | 65°C                  | alkaline | 100.00                  | 0.88                     | GCA_002760415    | Olsen et al., 2015            |
| *Thermostichus* sp. 65AY6Li    | Mushroom Spring, Yellowstone National Park, United States | 65°C                  | alkaline | 99.61                   | 1.32                     | GCA_002760345    | Olsen et al., 2015            |
| *Thermostichus* sp. JA-2-3Ba   | Octopus Spring, Yellowstone National Park, United States | 58–66°C               | 8.5      | 100.00                  | 0.00                     | GCA_000013225    | Bhaya et al., 2007            |
| *Thermostichus* sp. JA-3-3Ab   | Octopus Spring, Yellowstone National Park, United States | 58–66°C               | 8.5      | 100.00                  | 1.32                     | GCA_000013205    | Bhaya et al., 2007            |
| *Thermosynechococcus lividus* PCC 6715 | Hot spring, Yellowstone National Park, United States | NA                    | NA       | 98.86                   | 0.12                     | GCA_000013205    | This study                    |
| *Thermosynechococcus* sp. CL-1 | Chin-Lun hot spring, Taiwan, China       | 62°C                  | 9.3      | 100.00                  | 0.12                     | GCA_008386235    | Cheng et al., 2020            |
| *Thermosynechococcus* sp. TA-1 | Taian hot springs, Taiwan, China         | 50°C                  | 7–9      | 100.00                  | 0.12                     | GCA_017086385    | Leu et al., 2013              |
| *Thermosynechococcus vestitus* BP-1 | Hot spring, Beppu, Japan     | 52°C                  | 7.5      | 99.76                   | 0.12                     | GCA_000011345    | Nakamura et al., 2002         |
| *Thermosynechococcus vestitus* ES42 | Hot spring, Lotus lake, Sichuan, China | 67.2°C               | 7.96     | 100.00                  | 0.12                     | GCA_003555505    | Liang et al., 2019            |
| *Thermosynechococcus vulcanus* NIES-2134 | Yunomine Hot Spring, Wakayama, Japan | 50–57°C               | NA       | 99.76                   | 0.12                     | GCA_003990665    | Liang et al., 2018            |
Genome Sequencing and de novo Assembly

Additional to the 16 downloaded genomes, we sequenced the genome of another thermophilic cyanobacterium, purchased from Pasteur Culture Collection of Cyanobacteria as Synechococcus lividus PCC 6715. Whole-genome sequencing was carried out using integrated sequencing strategies, including PacBio and Illumina HiSeq Technology. The PacBio SMRT sequencing of PCC 6715 yielded 96,587 adapter-trimmed reads (subreads) with an average read length of approximately 11 kb, corresponding to 400-fold coverage. De novo assembly was performed using the hierarchical genome assembly process (HGAP) method implemented in SMRT analysis v2.3.0 (Chin et al., 2013), generating two contigs. The Illumina sequencing of PCC 6715 generated 5,566,668 filtered paired-end reads (clean data), providing approximately 520-fold coverage of the genome. The clean data of short reads generated by Illumina were assembled into contigs using SOAPdenovo v2.04 (Luo et al., 2012) with default parameters. Based on the contigs from SOAPdenovo assembler, the contigs derived from the HGAP method were comparatively examined to determine their continuity and concatenated into one closed circular chromosome. The genome obtained was mapped with Illumina reads using BWA v0.7.17 (Li and Durbin, 2009) and then Pilon v1.23 (Walker et al., 2014) to correct any assembly and sequence errors. The genome of PCC 6715 was annotated as a reference protein set. Based on the experimentally confirmed protein sequences of Synechococcus lividus sp. PCC 6803 were downloaded from the NCBI under the accession number CP018092.

Identification of Orthologous Proteins

Amino acid sequences of 28 proteins involved in CCM of Synechocystis sp. PCC 6803 were downloaded from the CyanoBase as a reference protein set. Based on the bidirectional best (BBH) criterion (Brilli et al., 2010), orthologous proteins involved in CCM of the studied species were identified using BLASTP with the following thresholds: E-value cut-off of 1E-6, ≥30% identity and 70% coverage. The identified proteins were homologous to these proteins in Synechocystis PCC 6803, namely bica1 (sll0834), bica2 (sll0096), cca1 (sll1347), ccmk1 (sll1029), ccmk2 (sll1028), ccmk3 (sll1388), ccmk4 (sll1389), ccmm (sll1030), ccmm (sll1031), ccmn (sll1032), ccmo (sll0436), ccmp (sll0169), cmp (sll040), cmpB (sll0041), cmpC (sll0043), cump (sll1373), cump (sll1030), ndhd3 (sll1733), ndhd4 (sll0027), ndhd4 (sll1732), ndhdf4 (sll0026), rbcL (sll0009), rbcS (sll0012), rbcX (sll0011), sbtA (sll1512), and sbtB (sll1513). The homologs of ecaA protein were searched using the experimentally confirmed protein ecaA (sll2929) of Anabaena PCC 7120 (Soltes-Rak et al., 1997) as a query. The italic number in bracket mentioned above refers to the gene ID in CyanoBase (see text footnote 2). The accession numbers of orthologous proteins identified in each genome of the thermophilic cyanobacteria studied was summarized in Supplementary Table 2.

Phylogeny

Amino acid sequences of Rubisco large subunit (rbcL) were collected for the studied species and reference cyanobacteria. The phylogram of rbcL was used to infer the protein function and classification among the strains. Protein sequences of genes ccmK, -L, -M, -N, -O, and -P encoding carboxysome shell proteins, cmpA, -B, -C, and -D encoding the HCO3 transporter BCT1 and nrtA, -B, -C, and -D of the nitrite/nitrate assimilation were also collected for phylogenetic reconstruction. All the protein sequences used were retrieved from the public databases, Cyanobase (see text footnote 2) or the NCBI3. All the phylogenetic analyses were performed using the following pipeline: multiple sequence alignments by Muscle complemented in Mega7 (Kumar et al., 2016) and Maximum-Likelihood (ML) inference of phylogenies by PhyML v3.3 (Guindon et al., 2010). Parameter settings in PhyML and bootstrap analysis of phylogenies were followed as described (Tang et al., 2019).

RESULTS AND DISCUSSION

Phylogenetic Analyses of Thermophilic Cyanobacteria

The 17 thermophilic cyanobacteria collected in this study all originated from hot springs (Table 1). Although the niche temperatures showed significant variation, these thermophilic cyanobacteria were defined based on their ability to grow above 50°C. Two types of morphology were characterized for these thermophilic cyanobacteria, comprising unicellular Synechococcus, Thermotrichus and Thermosynechococcus, and filamentous Leptodesmis, Leptolyngbya, Leptothrix, and Thermoleptolyngbya. In addition, strains JA-2-3Ba and JA-3-3Ab, previously known as Synechococcus strains, have been reclassified into a newly delineated taxon, genus Thermotrichus (Komárek et al., 2020). According to the suggested values for species (ANI > 96%, AAI ≥ 95%) and genus (ANI < 83%, AAI ≤ 70%) delimitation (Jain and Rodriguez, 2018), the result of genome-wide ANI and AAI indicated that strain JA-2-3Ba and JA-3-3Ab were different species (ANI = 85.6%, AAI = 87.8%) within Thermotrichus, while strain JA-3-3Ab, plus the other four strains from Yellowstone National Park (Table 1), belong to the same species (ANI/AAI > 98%) also within Thermotrichus. However, the case of S. lividus PCC 6715 was contradictory. Between S. lividus PCC 6715 and the five Thermosynechococcus strains, the values of ANI and ANI were ~ 77 and ~ 81% (Supplementary Table 3), respectively. To resolve the dilemma, a third approach was employed to determine the percentages of conserved proteins (POCP) between genomes. The POCP values (Supplementary Table 3) between S. lividus PCC 6715 and the Thermosynechococcus strains ranged from 83.4 to 84.0%, all far beyond the threshold (50%) for the definition of a prokaryotic genus ( Qin et al., 2014). Taken together, strain PCC 6715 might be a member of Thermosynechococcus, and was proposed

---

1http://genome.kazusa.or.jp/cyanobase

2https://www.ncbi.nlm.nih.gov/
to be *Thermosynechococcus lividus*. However, this conclusion requires validation with comprehensive investigation beyond the genomic data.

The carboxysome type of the 17 thermophilic cyanobacteria was investigated based on the phylogenetic analysis of a frequently used molecular marker, *rbcL* (Klanchui et al., 2017). The ML phylogram of *rbcL* categorized the 35 cyanobacteria into two categories (Figure 2), *α*-cyanobacteria and *β*-cyanobacteria, according to their Rubisco forms. There was only one exception that *Synechococcus* PCC 7942 was solely placed in a separate branch. The 17 thermophilic cyanobacteria surveyed were all grouped into the *β*-cyanobacteria category, indicating the presence of Rubisco 1B form in these thermophiles. However, considerable genetic diversity of *rbcL* amino acid sequences was also shown among these thermophilic cyanobacteria, distributing the strains into two clades (Figure 2). Clade A included only *Thermochromatium* strains, also known as early-branching or early-divergent *Synechococcus* (Blank and Sánchez-Baracaldo, 2010), whereas clade B contained thermophiles from a different genus. In addition, the clustering of strains did not completely comply with morphology, e.g., filamentous strains clustered with unicellular strains (clade B). Intriguingly, the 17 thermophilic strains did not group with any non-thermophilic strains, suggesting *rbcL* specificity of thermophiles. Although phylogenetic analysis of *rbcL* sequences can be used for the classification of cyanobacterial types, the evolutionary relationship based on habitats or morphology cannot be elucidated from the present *rbcL* phylogram. This is consistent
with the conclusions in previous studies that phylogenetic inference of molecular markers may be improper for the elaboration of evolutionary relationship with cyanobacterial habitat and environments (Komárek, 2016; Klanchui et al., 2017).

Genes Encoding C_i Uptake Systems in Thermophilic Cyanobacteria

Five different transport systems (Table 2) have been identified in the 17 thermophilic cyanobacteria. The two CO_2 uptake systems, NDH-1_3 and NDH-1_4 complex, were present in all the surveyed thermophilic cyanobacteria. Both CO_2 uptake systems in these thermophilic cyanobacteria were consistent with the previous reports that both CO_2 transporters were present in β-cyanobacteria living in freshwater, brackish or eutrophic lakes (Sandrini et al., 2014; Durall and Lindblad, 2015). In contrast, oceanic α-cyanobacteria (e.g., Prochlorococcus species) and marine β-cyanobacteria (e.g., Trichodesmium species) contained only one or even lacked them entirely (Price et al., 2007; Pronina et al., 2017). These results strongly indicated that the presence of NDH-1_3 and NDH-1_4 might be relevant to the environments where these cyanobacteria live. Thermophilic cyanobacteria, similar to freshwater and estuarine strains, contained essential components of CO_2 uptake and bicarbonate transport genes. Both constitutive systems NDH-1_4 complex and bicA were ubiquitous in all known thermophilic strains. The inducible NDH-1_3 complex is also universal suggesting that these strains are prepared for occasional carbon dioxide shortages. Unfortunately, it was not verified based on the current data that the niche temperature directly affected the existence of the two CO_2 uptake systems in light of their near-ubiquity in non-thermophilic cyanobacteria. The protein sequences of five genes encoding the two systems showed different identities with the sequences of non-thermophilic reference cyanobacteria (Synechocystis PCC 6803), ranging from 50.4 to 64.7% (Table 2), but a high degree of homology (≥70% identity) was noticed only among cupA genes. In addition, intragenus and intergenic variations of protein sequences were revealed by the different identities (Table 2). Moreover, the distinct property of the two complexes, a low-CO_2 inducible high-affinity CO_2 uptake system (NDH-1_3 complex) and a constitutive low-affinity CO_2 uptake system (NDH-1_4 complex), may confer the thermophilic cyanobacteria with more alternative strategies to survive in environments with significant CO_2 fluctuation, particularly in hot springs.

More variations among the 17 thermophilic cyanobacteria were noticed in HCO_3− transport systems than in CO_2 uptake systems (Table 2 and Figure 3). Association of the abundance of these mechanisms with their respective environmental niches is inconclusive. Similar repertoires of CO_2 and bicarbonate uptake genes were observed in L. sichuanensis A121 with the lowest habitat temperature and the most thermophilic strains of genus Thermostichus (Table 2 and Figure 3). When it comes to the pH of the environmental niches, no evident niches preference of genes was observed. Genetic repertoire of strains isolated from acidic and alkaline niches appeared to be unrelated to the pH of the environmental niches, no evident niches preference of genes was observed. Genetic repertoire of strains isolated from acidic and alkaline niches appeared to be unrelated to the pH of the environmental niches, no evident niches preference of genes was observed.

Table 2: Availability of genes encoding C_i uptake systems of thermophilic cyanobacteria studied.

| Strain                  | CO_2 uptake systems | HCO_3− transport systems |
|-------------------------|---------------------|--------------------------|
|                         | NDH-1_4 complex     | NDH-1_3 complex          | BicA | Sbt regulator | BCT1 |
|                         | ndhD4   | ndhF4   | cupB | ndhD3   | ndhF3   | cupA | bica1 | sbtA | sbtB | cmpA | cmpB | cmpC | cmpD |
| Leptolyngbya JC-1       | (80.9) | (59.7) | (62.7) | (81.5) | (57.2) | (71.9) | (68.5) | x | x | x | √ | (69.2) | (70.9) | (73.6) | (67.2) |
| Thermostichus E412      | (86.2) | (56.9) | (62.9) | (59.6) | (59.0) | (81.7) | (57.3) | x | √ | (75.9) | (79.1) | (63.9) | (65.2) | (70.8) | (72.2) |
| Thermoleptolyngbya A183 | (83.8) | (59.3) | (64.7) | (62.8) | (60.8) | (84.2) | (64.8) | (59.2) | x | x | √ | (63.0) | (64.5) | (70.8) | (73.6) |
| Thermoleptolyngbya O-77 | (80.3) | (55.6) | (62.3) | (58.8) | (57.8) | (85.7) | (65.5) | (62.3) | √ | (70.7) | (82.5) | x | x | x | x |
| Thermostichus 60AY6Li   | (75.0) | (54.5) | (61.3) | (57.7) | (67.7) | (84.7) | (65.5) | x | √ | (71.9) | (83.4) | (68.7) | (64.8) | (72.8) | (75.8) |
| Thermotrichus 60AY6S   | (57.4) | (54.5) | (56.8) | (58.8) | (53.9) | (75.8) | (80.2) | x | x | x | √ | (65.8) | (66.2) | (72.1) | (72.2) |
| Thermotrichus JA-28Ba  | (57.4) | (54.5) | (56.8) | (58.8) | (53.9) | (75.8) | (58.9) | x | x | x | √ | (65.8) | (66.2) | (72.2) | (66.8) |
| Thermotrichus JA-3AB   | (57.4) | (54.5) | (56.8) | (58.8) | (53.9) | (75.9) | (80.2) | x | x | x | √ | (65.4) | (66.2) | (72.2) | (67.1) |
| Thermosynechococcus PCC 6715 | (56.8) | (55.5) | (59.7) | (59.2) | (50.4) | (70.7) | (62.0) | x | x | x | √ | (67.5) | (66.2) | (71.8) | (65.0) |
| Thermostichus CL-1      | (59.3) | (56.7) | (61.9) | (61.5) | (58.1) | (70.6) | (57.2) | (74.1) | (80.9) | (66.5) | (59.5) | (73.1) | (71.2) |
| Thermotrichus TA-1      | (80.2) | (56.8) | (62.6) | (61.1) | (58.0) | (70.8) | (56.8) | x | √ | (82.0) | x | x | x | x |
| Thermosynechococcus BP-1 | (80.3) | (56.8) | (61.4) | (61.4) | (57.7) | (70.4) | (56.1) | x | x | √ | (67.7) | (61.0) | (73.3) | (66.8) |
| Thermotrichus ES42      | (59.9) | (57.4) | (62.0) | (62.9) | (59.0) | (70.4) | (56.4) | (72.7) | (81.8) | x | x | x | x |
| Thermosynechococcus NIES-2134 | (59.9) | (56.8) | (61.4) | (61.6) | (58.3) | (70.7) | (53.2) | x | x | √ | (67.5) | (68.1) | (73.0) | (68.1) |

√ and x refers to the presence and absence of the gene, respectively. The number in brackets indicates the identity (%) between the identified protein and the corresponding sequence of Synechocystis PCC 6803.
the other strains contained \textit{bicA1} or \textit{bicA2} (Table 2). The amino acid sequence identities of \textit{bicA} genes were around 60% between the thermophilic cyanobacteria and \textit{Synechocystis} PCC 6803. The presence of \textit{bicA1} and/or \textit{bicA2} was consistent within each genus except for \textit{Thermoleptolyngbya} (strain A183 and O-77). Sequence analysis suggested that a putative protein-coding gene of O-77 (O-77.5066, Supplementary Table 1) showed very low amino acid coverage (46.9%) to \textit{bicA2} of A183, but the aligned region was almost identical to each other (Supplementary Figure 1), indicating this short CDS might be a partial region of \textit{bicA2}. Such discrepancy in CDS length of \textit{bicA2} within \textit{Thermoleptolyngbya} requires future studies to elucidate the possible causes, e.g., mutations, gene lost, or sequencing errors. For \textit{sbt} regulator, both \textit{sbtA} and \textit{sbtB} were found only in six thermophilic cyanobacteria (Table 2). Furthermore, considerable variations existed within the genus \textit{Thermosynechococcus}. Strain E542, PCC 6715 and CL-1 possessed both \textit{sbtA} and \textit{sbtB}. TA-1 had only \textit{sbtB}, while the other two \textit{Thermosynechococcus} strains contained none. This result implied that the \textit{sbt} genes of these \textit{Thermosynechococcus} strains might be acquired by horizontal gene transfer or vertically inherited but lost during the evolutionary process. In addition, it appeared that the thermophilic cyanobacteria typically have \textit{bicA} rather than \textit{sbt} as revealed by the dominant presence (Table 2). This could be ascribed to the distinct traits of the two transporters and the habitat of these cyanobacteria. First, \textit{bicA} shows low affinity and high flux rate to HCO$_3^-$ and the genes encoding \textit{bicA} were found to be primarily constitutively expressed (Price et al., 2004). Second, \textit{sbt} was highly inducible under carbon-limited conditions and had a relatively high affinity to HCO$_3^-$ (Price, 2011). The \textit{sbt} may therefore become accessory when cyanobacteria live in alkaline ecosystems typically rich in HCO$_3^-$, which was in accordance with most cases in this study (Table 1), e.g., \textit{Thermostichus} strains. Meanwhile, it was evident that thermophilic cyanobacteria with both transporters might be advantageous to acclimatize to the highly dynamic shift of exogenous HCO$_3^-$ by selective utilization of the two HCO$_3^-$ transport systems.

As for the third HCO$_3^-$ transporter, BCT1 encoded by \textit{cmpA}, -B, -C, and -D and with high affinity for HCO$_3^-$ was induced under low levels of C$_i$ and enhanced by high light conditions (Omata et al., 2002). The \textit{cmp} genes were detected in all thermophilic cyanobacteria except for \textit{Thermoleptolyngbya} A183 and \textit{Thermosynechococcus} E542, PCC 6715 and TA-1 (Table 2). Intriguingly, strains without \textit{cmp} genes all possessed \textit{sbt}, which might to some extent compensate the function of \textit{cmp} genes for adaptation to thermal environments limited in inorganic carbon (Kamennaya et al., 2012). The identified \textit{cmp} genes showed amino acid sequence identity of 59.5–75.8% with the reference proteins (Table 2). In addition, it was reported that \textit{nrtA}, -B, -C, and -D (nitrate/nitrite transport system) shared high sequence similarity.
with cmpA, -B, -C, and -D (Omata, 1995). Therefore, we also identified the nrt genes in the 17 thermophilic cyanobacteria, and phylogenetic analysis based on amino acid sequences of cmp and nrt genes was performed to verify the annotations with the help of experimentally confirmed cmp proteins of Synechococcus PCC 7942 (Omata et al., 1999). As shown in Figure 4, the cmp proteins of thermophilic cyanobacteria clustered with that of Synechococcus PCC 7942 and other reference strains, while the nrt proteins grouped into clades or clusters separated from cmp proteins. Furthermore, sequence conservation was low, as suggested by long branches among the clusters formed in the phylogram (Figure 4). Surprisingly, the genome of Thermoleptolyngbya O-77 was the only genome without the complete set of nrtA, -B, -C, and -D, nrtC of which was not found by BLASTP. However, sequence analysis (Supplementary Figure 2) showed that two putative genes both shared high similarity of protein sequence with nrtC of Thermoleptolyngbya A183, suggesting unknown interruption within CDS of nrtC in O-77 genome. This might be ascribed to mutations, partial gene lost, or sequencing errors. Intriguingly, the discrepancy of the presence of cmp genes within genus Thermoleptolyngbya and Thermosynechococcus again suggested that these genes might be acquired by horizontal gene transfer or vertically inherited but lost during the evolutionary process.

In summary, the C uptake systems could assign the 17 thermophilic cyanobacteria into three genotypes: (I) strains comprising NDH-1, NDH-1, bicA, sbt, and BCT1; (II) NDH-1, NDH-1, bicA, and sbt; and (III) NDH-1, NDH-1, bicA, and BCT1. This result suggested that the same CO2 uptake systems and different HCO3− transport systems were shared by these thermophilic cyanobacteria. Different genotypes were primarily attributed to the presence/absence of sbt and BCT1 (Table 2). Only three strains, namely Leptolyngbya JSC-1, Thermoleptolyngbya O-77, and Thermosynechococcus CL-1, contained both types of high-affinity HCO3− transporters, indicating that these strains may have a higher capacity for HCO3− uptake under various growth environments. Definitely, owning one of sbt and BCT1 appeared to be sufficient for thermophilic strains to cope with extremely low inorganic carbon and ions in thermal aquatic environments. Future experimental studies should be carried out to elucidate the function of these in silico identified proteins in thermophilic cyanobacteria.

### Genes Encoding Carboxysomes in Thermophilic Cyanobacteria

Carboxysomes in cyanobacteria are bacterial microcompartments (BMCs) that provide an environment for enhancing the catalytic capabilities of the CO2 fixing enzyme, Rubisco (Cai et al., 2009). A typical β-cyanobacterial genome usually contains ccmK1-4 (MacCready et al., 2020). However, only three out of 17 thermophilic cyanobacteria studied showed such diversity of the shell proteins, while Thermoleptolyngbya strains lacked ccmk2, and Thermotrichus and Thermosynechococcus strains contained none of ccmk3 and ccmK4 (Table 3). Previous studies showed that ccmK1 and ccmK2 were the main structural proteins of the carboxysome shell (Kerfeld et al., 2005; Rae et al., 2013). Phylogenetic analysis and identity calculation suggested particularly high sequence conservation between the two homologs of ccmK1/2, as revealed by low amino acid substitution rates and short branches (Figure 5A) and high identity (Table 3). In contrast, phylogenetic branches of the ccmK3 and ccmK4 proteins were much longer, and residues were less conserved (Figures 5B,C and Table 3). Although deletion of the ccmK2 gene in Synechococcus PCC 7942 that lacked ccmK1 caused a carboxysome-less phenotype (Rae et al., 2012), this experiment did not verify whether normal phenotype can be achieved with the presence of sole ccmK1 homolog. Besides, the main distinguishing feature between ccmK1 and ccmK2 was the ~10 amino acid long C-terminal extension of ccmK1 (Kerfeld et al., 2005), which has been proven to show limited structural relevance for shell assembly (Cai et al., 2016). Instead, it might be just the genomic proximity of the two ccmK genes, as indicated by its high degree of conservation (15/17 genomes, Table 3). Future studies are required to carefully investigate the specificity regarding the structure and function of ccmK1 in the Thermoleptolyngbya strains. Regarding ccmK3 and ccmK4, the widespread absence of the two genes was not surprising. The previous studies showed they are similar molecular components (Sommer et al., 2017) and suggested functional redundancy that play only a minor structural role (Rae et al., 2012). On the other hand, a recent study proposed a functional scenario in which minor ccmK3/K4 incorporation in shells would introduce sufficient local disorder to allow shell remodeling (Garcia-Alles et al., 2019). Another study indicated that ccmK3/ccmK4 heterohexamers potentially expand the range of permeability properties of metabolite channels in carboxysome shells (Sommer et al., 2019). Taken together, these findings suggested that ccmK3/K4 present in Leptodesmis A121, Leptolyngbya JSC-1, Leptothermofonsea E412, and Thermoleptolyngbya A183 and O-77 may be utilized to adjust the properties of carboxysome for rapid adaptation to environmental changes that happen in thermal regions. In addition to ccmK, other genes encoding carboxysome shell proteins, ccmL, -M, -N, -O, and -P were also found to be present in all 17 thermophilic cyanobacteria (Table 3). Only ccmN exhibited weak homologs (<40%). All the surveyed thermophilic cyanobacteria showed a high degree of homology (67.6~76.1%) with the ccmP of Synechocystis PCC 6803, and similarly identity range of 65.8~75.5% to the ccmP of Synechococcus PCC 7942, an experimentally confirmed protein (Cai et al., 2013). Further phylogenetic analysis suggested extensive genetic diversity in these genes as indicated by the assignments of these thermophilic cyanobacteria into different clusters or clades (Supplementary Figure 3).

The subunits of Rubisco, rbcL and rbcS, and Rubisco assembly chaperone, rbcX, were present in all the 17 thermophilic cyanobacteria (Table 3). The protein sequences of rbcL in thermophilic cyanobacteria were highly conserved to the reference protein (>83% identity), while that of rbcS was moderately conserved (~65% identity). The protein sequences of rbcX were more variable, showing ~45% identity with the reference protein. In regards to β-CA, carboxysomal ccaA and/or non-carboxysomal ecaB proteins with low sequence identity (38.1~49.7%) were detected only in six thermophilic
FIGURE 4 | Phylogenetic inference of protein sequences of cmp and nrt genes. (A) cmpA and nrtA; (B) cmpB and nrtB; (C) cmpC and nrtC; (D) cmpD and nrtD. The thermophilic cyanobacteria investigated in this study are indicated in bold. The accession numbers underscored refers to the gene IDs in Supplementary Table 1. Only bootstrap values > 50% are indicated at nodes.
strains (Table 3), respectively. Apart from the functions as shell protein, an additional function of ccmM (γ-CA) as CA activity was reported in cyanobacteria lacking ccaA (Peña et al., 2010; de Araujo et al., 2014). Thus, we compared γ-CA-like domain in N-terminal of ccmM in thermophilic cyanobacteria to the functional γ-CA in Thermosynechococcus BP-1 (Peña et al., 2010) and Nostoc PCC 7120 (de Araujo et al., 2014) as well as non-functional γ-CA in Synechocystis PCC 6803 (Cot et al., 2008) and Synechococcus PCC 7942 (So and Espie, 2005). As shown in Figure 6, strains from Thermotrichus and Thermosynechococcus exhibited similar primary structure of amino acid residues in γ-CA-like domain of ccmM to that with CA activity in Thermosynechococcus BP-1 and Nostoc PCC 7120. The result suggested that these ccmM proteins may play a role in CA activity in light of the absence of a carboxysomal CA (ccaA) (Table 3) in these thermophilic cyanobacteria. Although strains from Leptodesmis, Leptolyngbya, Leptothermolamia, and Thermoleptolyngbya had both ccmM and ccaA, it is likely that only ccaA function as CA activity as indicated by the primary structure analysis, which was consistent with the result found in haloalkaliphilic cyanobacteria Microcoleus sp. IPPAS B-353 (Kupriyanova et al., 2016). In addition, strains living in high temperatures above 50°C (Table 1) appeared to tend to lose ccaA. Meanwhile, high pH conditions were speculated to be an important factor in ccaA absence (Klanchui et al., 2017). However, the reasons were still unclear. Certainly, the co-evolution of ccmM and ccaA is crucial for carboxysome to regularly function in extreme environments regardless of high temperature or alkalinity. As for α-CA, non-carboxysomal ecaA was not found in all the 17 thermophilic cyanobacteria.

**Genomic Distribution Pattern of Carbon-Concentrating Mechanism-Related Genes in Thermophilic Cyanobacteria**

The genomic location of CCM-related genes in thermophilic cyanobacteria may provide insights into the function and evolution of these genes. In the present study, the genomic organization of CCM-related genes was illustrated in Figure 7. Genes encoding each NDH-1 complex were all clustered together, except for the ndhD4 of Thermosynechococcus TA-1. This gene was remote from the other genes of NDH-1 complex (Figure 7). In light of the coordination of NDH and cup proteins in CO₂
TABLE 3 | Availability of genes encoding carboxysomes of thermophilic cyanobacteria studied.

| Strain                        | Carboxysome | Shell proteins | Encapsulated enzymes | Rubisco | β-CA |
|------------------------------|-------------|----------------|----------------------|---------|------|
|                              |             | β-Carbonosomal shell proteins |                      |         |      |
|                              | ccmK1       | ccmK2          | ccmK3                | ccmK4   | ccmL | ccmM | ccmN   | ccmO  | ccmP   | rbcL | rbcS | rbcX | ccaA | ecaB |
| Leptothecosmis A121          | ✓           | ✓              | ✓                    | ✓       | ✓    | ✓    | ✓      | ✓     | ✓      | ✓    | ✓    | ✓    | ✓    | ✓    |
| Leptolyngbya JSC-1           | ✓           | ✓              | ✓                    | ✓       | ✓    | ✓    | ✓      | ✓     | ✓      | ✓    | ✓    | ✓    | ✓    | ✓    |
| Leptothecosmis E412          | ✓           | ✓              | ✓                    | ✓       | ✓    | ✓    | ✓      | ✓     | ✓      | ✓    | ✓    | ✓    | ✓    | ✓    |
| Thermosynechococcus A183     | ✓           | ✓              | ✓                    | ✓       | ✓    | ✓    | ✓      | ✓     | ✓      | ✓    | ✓    | ✓    | ✓    | ✓    |
| Thermosynechococcus O-77     | ✓           | ✓              | ✓                    | ✓       | ✓    | ✓    | ✓      | ✓     | ✓      | ✓    | ✓    | ✓    | ✓    | ✓    |
| Thermostichus 60AY4M2        | ✓           | ✓              | ✓                    | ✓       | ✓    | ✓    | ✓      | ✓     | ✓      | ✓    | ✓    | ✓    | ✓    | ✓    |
| Thermostichus 63AY4M2        | ✓           | ✓              | ✓                    | ✓       | ✓    | ✓    | ✓      | ✓     | ✓      | ✓    | ✓    | ✓    | ✓    | ✓    |
| Thermostichus 65AY6A5        | ✓           | ✓              | ✓                    | ✓       | ✓    | ✓    | ✓      | ✓     | ✓      | ✓    | ✓    | ✓    | ✓    | ✓    |
| Thermostichus 65AY6LI        | ✓           | ✓              | ✓                    | ✓       | ✓    | ✓    | ✓      | ✓     | ✓      | ✓    | ✓    | ✓    | ✓    | ✓    |
| Thermostichus JA-2-3Ba       | ✓           | ✓              | ✓                    | ✓       | ✓    | ✓    | ✓      | ✓     | ✓      | ✓    | ✓    | ✓    | ✓    | ✓    |
| Thermostichus JA-3-3Ab       | ✓           | ✓              | ✓                    | ✓       | ✓    | ✓    | ✓      | ✓     | ✓      | ✓    | ✓    | ✓    | ✓    | ✓    |
| Thermosynechococcus PCC 6715 | ✓           | ✓              | ✓                    | ✓       | ✓    | ✓    | ✓      | ✓     | ✓      | ✓    | ✓    | ✓    | ✓    | ✓    |
| Thermosynechococcus CL-1     | ✓           | ✓              | ✓                    | ✓       | ✓    | ✓    | ✓      | ✓     | ✓      | ✓    | ✓    | ✓    | ✓    | ✓    |
| Thermosynechococcus TA-1     | ✓           | ✓              | ✓                    | ✓       | ✓    | ✓    | ✓      | ✓     | ✓      | ✓    | ✓    | ✓    | ✓    | ✓    |
| Thermosynechococcus BP-1     | ✓           | ✓              | ✓                    | ✓       | ✓    | ✓    | ✓      | ✓     | ✓      | ✓    | ✓    | ✓    | ✓    | ✓    |
| Thermosynechococcus E542     | ✓           | ✓              | ✓                    | ✓       | ✓    | ✓    | ✓      | ✓     | ✓      | ✓    | ✓    | ✓    | ✓    | ✓    |
| Thermosynechococcus NIES-2134| ✓           | ✓              | ✓                    | ✓       | ✓    | ✓    | ✓      | ✓     | ✓      | ✓    | ✓    | ✓    | ✓    | ✓    |
| NIES-2134                    | ✓           | ✓              | ✓                    | ✓       | ✓    | ✓    | ✓      | ✓     | ✓      | ✓    | ✓    | ✓    | ✓    | ✓    |

✓ and x refers to the presence and absence of the gene, respectively. The number in brackets indicates the identity (%) between the identified protein and the corresponding sequence of Synechocystis PCC 6803.

FIGURE 6 | Partial alignments of ccmM amino acid sequences representing 17 thermophilic cyanobacteria. Nostoc PCC 7120, and Synechococcus PCC 7942 as outgroup. As suggested by Peña et al. (2010), red boxes refer to conserved regions of the N-terminal domain of ccmM necessary for CA activity, while shaded cysteine amino acids indicate essential residues participating in the disulfide bond in the C-terminus of active ccmM protein.

uptake (Han et al., 2017), whether and how such organization in Thermosynechococcus TA-1 affects the co-regulation requires further experimental elucidation. In addition, the genes encoding bicA1 and bicA2 were positionally scattered in the genomes of Leptothecosmis E412 and Thermosynechococcus A183, whereas genes encoding cmpABCD and sbtA/B were located together in genomes (Figure 7), respectively.

Focusing on the gene organization of the carboxysome, there appeared to be the main carboxysome locus (MCL) in the genomes of all the studied thermophilic cyanobacteria, comprising one to two ccmK1/2 genes, always followed by ccmL, -M, -N and in the majority of genomes (11 out of 17), ccmO (Figure 7). The sequential arrangement of ccmK2 and ccmK1 in the MCL might facilitate protein complex assembly or help
to balance the shell protein stoichiometry during translation of MCL genes (Cai et al., 2016). Moreover, previous studies suggested that ccmK, -L, -M, and -N were co-regulated in an MCL as an operon (Rae et al., 2013; Billis et al., 2014). In addition to the MCL, satellite loci were found in several genomes. Satellite locus encoding the ccmP protein was found in all the 17 thermophilic cyanobacteria. The satellite loci formed by ccmK3 and ccmK4 was present only in five genomes (Figure 7). The organization of K1/K2 and K3/K4 paralogs in separated operons might relate to structural segregation of the two groups (Sommer et al., 2019). Although ccmK3 and ccmK4 clustered in operon, it was reported that the stoichiometry of incorporation of ccmK3 in associations with ccmK4 was low in Synechocystis PCC6803 (Garcia-Alles et al., 2019). However, the expression of ccmK3 and ccmK4 from a satellite locus may increase the flexibility of carboxysome shell assembly and permeability (Sommer et al., 2019), and may provide differing metabolite selectivities (Sommer et al., 2017). Another satellite locus, ccmO, was positionally variable (Figure 7). In 11 genomes, ccmO was the terminal gene in the MCL. However, ccmO in six Thermosynechococcus genomes was located as a satellite locus, remote from other carboxysome genes, and this ccmO formed a distinct clade as revealed by the phylogenetic analysis (Supplementary Figure 3D). Previous transcriptome analysis demonstrated that the ccmO gene was co-regulated with MCL genes only when it was associated with the MCL (Sommer et al., 2017). Therefore, the dispersal of ccmO as well as other carboxysome genes across the genome may increase the plasticity of individual regulation of their expression, perhaps in response to changes in environmental conditions, and vice versa, diverse environmental conditions appear to promote relocation of these genes driven by evolutionary force (Cai et al., 2012; Shih et al., 2013).

Genes encoding rbcL, rbcS, and rbcX for Rubisco were localized together in all of the 17 thermophilic cyanobacteria, which was in line with the previous consensus that the three genes were clustered in an operon in β-cyanobacteria (Badger et al., 2002). The Rubisco gene cluster was all apart from the MCL in each genome, indicating that it, as satellite loci to MCL, might regulate its expression independently. This was supported by the results reported by Billis et al. (2014) that ccmO, rbcL, and rbcS were less strictly co-regulated with the MCL genes. As for the chaperone rbcX, sometimes it was non-essential for the
assembly of an active Rubisco enzyme, e.g., in *Synechococcus* PCC 7942 (Emlyn-Jones et al., 2006). The role of *rbcX* in these thermophilic cyanobacteria requires future investigations. In regards to β-CA*s, scattered distribution of *ccaA* and *ecaB* was present in the genomes.

Functional and Comparative Analysis of Carbon-Concentrating Mechanism Components Among β-Cyanobacteria

The overall composition of the CCM system presented in Figure 3 contains all typical components found in other cyanobacteria, including the complete set of genes for carboxysome assembly. A comparison of these components among β-cyanobacteria was performed between thermophilic cyanobacteria and non-thermophilic cyanobacteria. For a precise comparison, we divided the non-thermophilic cyanobacteria into three groups: freshwater, marine, and alkaliophilic cyanobacteria. The overall compositions of CCM components in thermophilic cyanobacteria were more similar to the freshwater and alkaliophilic than the marine groups (Klanchui et al., 2017). The thermophilic, freshwater and alkaliophilic cyanobacteria possessed both CO$_2$ uptake systems, NDH-13 and NDH-11, whereas most marine cyanobacteria lacked the NDH-13. In light of the HCO$_3^-$ uptake, marine and some thermophilic/alkaliophilic cyanobacteria lacked the BCT1 transporter. Although freshwater cyanobacteria consistently possessed BCT1 transporter, some thermophilic cyanobacteria with freshwater origins lacked that. The relatively high diversity of bicarbonate transporters is a distinguishing factor for certain thermophilic strains probably as indicator of their specific location and not their thermophilic character. However, there is no apparent correlation between the pH of the isolation sources of these strains and the presence or absence of specific bicarbonate uptake systems (Figure 3 and Table 1). More thorough studies are required to characterize the functions of these transporters in the future. In addition, the freshwater cyanobacteria possessed the highest abundance of CAs, β-CA (*ccaA* and *ecaB*), α-CA (*ecaA*), and γ-CA (*ccmM*), while the thermophilic cyanobacteria with freshwater origins were likely to possess up to two CAs (Table 3 and Figure 3). Notably, 11 out of the 17 investigated thermophiles appeared to have only γ-CA (*ccmM*). On the other hand, the diversity of carboxysome shell proteins appears to be lower in thermophilic strains than many of their mesophilic counterparts, with multiple strains lacking at least one of the four *ccmK* genes.

The distinct genomic distribution of CCM-related genes among the thermophiles probably suggested various regulations of expression. Overall, the comparative genomic analysis revealed distinct molecular components and organization of CCM in thermophilic cyanobacteria. These findings provided insights into the CCM components of thermophilic cyanobacteria and fundamental knowledge for further research regarding photosynthetic improvement and biomass yield of thermophilic cyanobacteria with biotechnological potentials.

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: https://www.ncbi.nlm.nih.gov/, CP018092.

AUTHOR CONTRIBUTIONS

JT: conceptualization, methodology, validation, formal analysis, investigation, data curation, writing – original draft, writing, review, editing, visualization, supervision, project administration, and funding acquisition. HZ: formal analysis, investigation, data curation, and visualization. DY: formal analysis, investigation, and data curation. SR: formal analysis, data curation, and visualization. AK-S: formal analysis, visualization, writing, review, and editing. AK-S: formal analysis, investigation, and funding acquisition. MD: conceptualization, methodology, resources, data curation, writing – original draft, writing, review, editing, supervision, project administration, and funding acquisition. All authors contributed to the article and approved the submitted version.

FUNDING

This research was funded by the National Natural Science Foundation of China (31970092 and 32071480), Shenzhen Fundamental Research Program (GXWD20201231165807007-20200806170221001), National Science Centre (Poland) through project no. 2018/31/B/ST8/00822, and Tenure-Track Fund to MD. Funding bodies had no influence over the design and execution of this research.

ACKNOWLEDGMENTS

We would like to thank Dr. Igor Brown and the culture collections of PCC and NIES for the assistance in verifying the actual thermophilicity of certain strains.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fmicb.2022.876272/full#supplementary-material
REFERENCES

Alcorta, J., Alarcón-Schumacher, T., Salgado, O., and Diez, B. (2020). Taxonomic novelty and distinctive genomic features of hot spring cyanobacteria. *Front. Genet.* 11, 568223. doi: 10.3389/fgene.2020.568223

Aziz, R. K., Bartels, D., Best, A. A., Dejongh, M., Disz, T., Edwards, R. A., et al. (2008). The RAST Server: Rapid annotations using subsystems technology. * BMC Genom.* 9:75. doi: 10.1186/1471-2164-9-75

Badger, M. R., Hanson, D., and Price, G. D. (2002). Evolution and diversity of CO2 concentrating mechanisms in cyanobacteria. *Funct. Plant Biol.* 29, 161–173. doi: 10.1071/FP001213

Bhaya, D., Grossman, A. R., Steunou, A.-S., Khuri, N., Cohan, F. M., Hamamura, N., et al. (2007). Population level functional diversity in a microbial community revealed by comparative genomic and metagenomic analyses. *ISME J.* 1, 703–713. doi: 10.1038/ismej.2007.46

Billis, K., Billini, M., Tripp, H. J., Kyrpides, N. C., and Mavromatis, K. (2020). Taxonomic novelty and distinctive genomic features of hot spring cyanobacteria. *Front. Microbiol.* 11, 263–270. doi: 10.3389/fmicb.2015.00702

Emlyn-Jones, D., Woodger, F. J., and Whitney, P. S. M. (2006). RbcX can function as a rubisco cheropunin, but is non-essential in *Synechococcus* PCC7942. *Plant Cell Physiol.* 47, 1630–1640. doi: 10.1093/pcp/pcl028

Galmés, J., Hermida Carrera, C., Laaistio, L., and Ninemets, U. (2016). A compendium of temperature responses of Rubisco kinetic traits: variability among and within photosynthetic groups and impacts on photosynthesis modeling. *J. Exp. Bot.* 67, erw267. doi: 10.1093/jxb/erw267

Hénmans, J., Kapralov, M. V., Copolovici, L. O., Hermida-Carrera, C., and Ninemets, U. (2015). Temperature responses of the Rubisco maximum carboxylase activity across domains of life: phylogenetic signals, trade-offs, and importance for carbon gain. *Photosyn. Res.* 123, 183–201. doi: 10.1007/s11110-014-0607-8

García-Alles, L., Root, K., Maveyraud, L., Aubry, N., Lesniewska, E., Mourey, L., et al. (2019). Occurrence and stability of hetero-hexamer associations formed by β-carboxysome CmCk shell components. *PLoS One* 9:e109738. doi: 10.1371/journal.pone.0109738

Blank, C. E., and Sánchez-Baracaldo, P. (2010). Timing of morphological and ecological innovations in the cyanobacteria—a key to understanding the rise in atmospheric oxygen. *Geobiology* 8, 1–23. doi: 10.1111/j.1472-4669.2009.00220.x

Brilli, M., Fondi, M., Fani, R., Mengoni, A., Ferri, L., Bazzicalupo, M., et al. (2010). The diversity and evolution of cell cycle regulation in alpha-proteobacteria: a comparative genomic analysis. *BMC Syst. Biol.* 4:52. doi: 10.1186/1752-0509-4-52

Brown, I. I., Bryant, D. A., Casamatta, D., Thomas-Keptra, K. L., Sarkisova, S. A., Shen, G., et al. (2010). Polyphasic characterization of a thermotolerant siderophilic filamentous cyanobacterium that produces intracellular iron deposits. *Appl. Environ. Microbiol.* 76, 6664–6672. doi: 10.1128/aem.00662-10

Burnap, R. L., Hagemann, M., and Kaplan, A. (2015). Regulation of CO2 concentrating mechanism in cyanobacteria. *Life S.* 5, 348–371. doi: 10.3390/life50100348

Cai, F., Bernstein, S. L., and Wilson, S. C. (2016). Production and characterization of synthetic carboxysome shells with incorporated luminal proteins. *Plant Physiol.* 170, 1868–1877. doi: 10.1104/pp.15.01822

Cai, F., Kerfeld, C. A., and Sandh, G. (2012). “Bioinformatic identification and structural characterization of a new carboxysome shell protein,” in *Functional Genomics and Evolution of Photosynthetic Systems*, eds R. Burnap and W. Vermaas (Dordrecth: Springer Netherlands), 345–356. doi: 10.1007/jbml.2009.03.056

Cai, F., Menon, B. B., Cannon, G. C., Curry, K. J., Shively, J. M., and Heinhorst, S. (2009). The pentameric vertex proteins are necessary for the icosahedral carboxysome shell to function as a CO2 concentrating mechanism. *PLoS One* 4:e7521. doi: 10.1371/journal.pone.0007521

Cai, F., Sutter, M., Cameron, J. C., Stanley, D. N., Kinney, J. N., and Kerfeld, C. A. (2013). The structure of CcmP, a tandem bacterial microcompartment domain protein from the β-Carboxysome, forms a subcompartment within a microcompartment. *J. Biol. Chem.* 288, 16055–16063. doi: 10.1074/jbc.M113.456987

Chen, H., and Wang, Q. (2020). Microalga-based nitrogen bioremediation. *Algal Res.* 46:101775. doi: 10.1016/j.algal.2019.101775

Chen, M.-Y., Teng, W.-K., Zhao, L., Hu, C.-X., Zhou, Y.-K., Han, B.-P., et al. (2021). Comparative genomic reveals insights into cyanobacterial evolution and habitat adaptation. *ISME J.* 15, 211–227. doi: 10.1038/s41396-020-00775-z

Cheng, Y.-L., Chou, L., Chiu, Y.-F., Hsieh, H.-T., Kuo, C.-H., and Chu, H.-A. (2020). Comparative genomic analysis of a novel strain of Taiwan hot-spring cyanobacterium *Thermosynechococcus* sp. CL-1. *Front. Microbiol.* 11:82. doi: 10.3389/fmicb.2020.00082

Chin, C.-S., Alexander, D. H., Marks, P., Klammer, A. A., Drake, J., Heiner, C., et al. (2013). Nonhybrid, finished microbial genome assemblies from long-read SMRT sequencing data. *Nat. Meth.* 10, 563–569. doi: 10.1038/nmeth.2474

Clement, R., Dimmet, L., Maberly, S. C., and Gonter, R. (2016). The nature of the CO2-concentrating mechanisms in a marine diatom, *Thalassiosira pseudonana*. *New Phytol.* 209, 1417–1427. doi: 10.1111/nph.13728

Cot, S. S. W., So, A. K. C., and Espie, G. S. (2008). A multipeptide bicarbonate dehydrogen complex essential to carboxysome function in cyanobacteria. *J. Bacteriol.* 190, 936–945. doi: 10.1128/jb.01283-07
Tang, J., Shah, M. R., Yao, D., Li, L., Zhao, K., Li, L., et al. (2022b). Polyphasic identification and genomic insights of *Leptothermofonsia sichuanensis* gen. sp. nov., a novel thermophilic cyanobacteria within Leptolyngbyaceae. *Front. Microbiol.* 13:765105. doi: 10.3389/fmicb.2022.765105

Tang, J., Du, L., Liang, Y., and Daroch, M. (2019). Complete genome sequence and comparative analysis of *Synechococcus* sp. CS-601 (SynAce01), a cold-adapted cyanobacterium from an oligotrophic Antarctic habitat. *Int. J. Mol. Sci.* 20:152. doi: 10.3390/ijms20010152

Tang, J., Jiang, D., Luo, Y., Liang, Y., Li, L., Shah, M., et al. (2018a). Potential new genera of cyanobacterial strains isolated from thermal springs of western Sichuan, China. *Algal Res.* 31, 14-20. doi: 10.1016/j.algal.2018.01.008

Tang, J., Liang, Y., Jiang, D., Li, L., Luo, Y., Shah, M. M. R., et al. (2018b). Temperature-controlled thermophilic bacterial communities in hot springs of western Sichuan, China. *BMC Microbiol.* 18:134. doi: 10.1186/s12866-018-1271-z

Tang, J., Li, L., Li, M., Du, L., Shah, M. R., Waleron, M., et al. (2021). Description, taxonomy, and comparative genomics of a novel species, *Thermoleptolyngbya sichuanensis* sp. nov., isolated from Hot Springs of Ganzi, Sichuan, China. *Front. Microbiol.* 12:696102. doi: 10.3389/fmicb.2021.696102

Tremaine, P., Zhang, K., Bénézeth, P., and Xiao, C. (2004). "Ionization equilibria of acids and bases under hydrothermal conditions," in *Aqueous Systems at Elevated Temperatures and Pressures*, eds D. A. Palmer, R. Fernández-Prini, and A. H. Harvey (London: Academic Press), 441-492. doi: 10.1016/b978-012544461-3/50014-4

Walker, B. J., Abeel, T., Shea, T., Priest, M., Abouelliel, A., Saktihikumar, 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

Walter, J. M., Coutinho, F. H., Dutílh, B. E., Swings, J., Thompson, F. L., and Thompson, C. C. (2017). Ecogenomics and taxonomy of cyanobacteria phylum. *Front. Microbiol.* 8:2132. doi: 10.3389/fmicb.2017.02132

Wooley, R. J. (2021). Evaluation of the temperature dependence of dissociation constants for the marine carbon system using pH and certified reference materials. *Mar. Chem.* 229:103914. doi: 10.1016/j.marchem.2020.103914

Yoon, K. S., Nguyen, N. T., Tran, K. T., Tsuji, K., and Ogo, S. (2017). Nitrogen fixation genes and nitrogenase activity of the non-heterocystous cyanobacterium *Thermoleptolyngbya* sp. O-77. *Microb. Environ.* 32, 324-329. doi: 10.1264/jsme2.ME17015

**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.

Copyright © 2022 Tang, Zhou, Yao, Riaz, You, Klepacz-Smółka and Daroch. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.