Posted on 10 Jan 2023 — The copyright holder is the author/funder. All rights reserved. No reuse without permission. — https://doi.org/10.22541/au.167398511.17288816/v1 — This a preprint and has not been peer reviewed. Data may be preliminary.

Draft genome sequence of a Spirulina subsalsa strain with high tolerance to salinity and hyperaccumulation of phycocyanin

Liqun Jiang[1], Hangzhou Xu[2], and Haiyan Pei[1]

[1]Fudan University Department of Environmental Science and Engineering
[2]Shandong University School of Environmental Science and Engineering Sciences

January 10, 2023

Abstract

The potential use of *Spirulina subsalsa* in marine cultivation is important for biomass and phycocyanin production; however, little is known about its mechanism of salt adaptation. Here we present a draft genome sequence for the halotolerant cyanobacterium *Spirulina subsalsa* FACHB-351 (5.3 Mbp). The sequence is predominantly distributed in signal transduction, cell growth and adaptation, and carbohydrate metabolism, which possibly allows the strong salt adaptation of FACHB-351. No genes directly related to bacterial toxins suggests a low risk of FACHB-351 threatening human or animal health. A set of genes annotated in the GO database were not assigned in the KEGG database, and some genes of essential enzymes, like HO and DVR for phycocyanin and chlorophyll, were not found, which suggests a presence of possible yet-to-be-discovered pathways in FACHB-351. The availability of a genome sequence will facilitate investigations into *S. subsalsa* phycocyanin biosynthesis and permit optimal strategies to improve commercial phycocyanin production.

1. Introduction

The ability to produce a heterogeneous group of bioactive compounds, including nutraceuticals and pharmaceuticals, makes algae an ideal vehicle for high-value products, and extra benefits from alga-based bioproducts are derived from the algae's survival in extreme conditions or competitive environments (Gorelova et al., 2019). Hence, taking into account the high cost of freshwater cultivation, wastewater and seawater have been commonly considered for alternative cultivation strategies. Non-competition with freshwater resources makes marine cultivation of algae an attractive platform for sustainable production of algal biomass, while also setting a high standard for strains based on the algal characteristics and the economic feasibility — most importantly the strain's tolerance to salinity (Pei and Jiang, 2018). In our previous studies on performance of strains with potential for industrial biomass production in seawater (Jiang et al., 2015; 2019), one of the winners was identified as the cyanobacterium *Spirulina subsalsa*FACHB-351, on account of its fast growth rate (0.3 d^{-1}), and strong adaptation to seawater-based medium. Additional advantages include its filamentous shape (length > 200 μm) for filtration harvest and high phycocyanin content for food additive production. Although *S. subsalsa*exhibited such competitive potential in terms of biomass production from marine cultivation, little is known about how it responds to environmental shifts.

Various salt tolerance mechanisms have been identified so far in cyanobacteria, including sodium-related membrane transporters, synthesis of compatible solutes, and salt-stress signalling (Cui et al., 2020). These mechanisms were mostly discovered in limnetic *Synechococcus*spp. and *Synechocystis* spp., due to their already-sequenced genome and the developed genetic toolbox. However, some salt resistant mechanisms are species-specific, and adaptation of *Spirulina subsalsa* FACHB-351 is promising to marine cultivation; therefore, it is necessary to understand its salt resistance mechanisms. Furthermore, halotolerant *Spirulina subsalsa* could be a better candidate in researching cyanobacterial response to the high salinity of seawater cultivation, due to their natural resistance to salt through evolutionary selection. However, so far the existing

Posted on 10 Jan 2023 — The copyright holder is the author/funder. All rights reserved. No reuse without permission. — https://doi.org/10.22541/au.167393851.17288816/v1 — This a preprint and has not been peer reviewed. Data may be preliminary.

knowledge of halotolerant cyanobacteria still significantly lags behind that for freshwater cyanobacteria. It can be addressed through a whole genome sequencing project, which lays the foundation of future metabolic pathway studies.

Herein, to make *S. subsalsa* FACHB-351 a model system for marine cultivation and research into cyanobacterial response to marine cultivation, this study determined the nearly complete genome sequence of the filamentous cyanobacterium FACHB-351, with a length of 5.3 Mbp. Genomic features were compared among strains that are morphologically and molecularly related to *S. subsalsa* , and some characteristic gene sets — particularly in signal transduction, photosynthetic pigment synthesis, carbohydrate generation, and cell motility — were determined. The FACHB-351 genome sequence and its curated annotation are important assets to better understand the physiology and metabolic potential of *Spirulina* and will open up new opportunities in the functional genomics of this species.

2. Results

2.1 Strong adaptation

The biomass yields of *S. subsalsa* FACHB-351 under the condition of extra NaCl addition are shown in Figure 1a. Together with previous observations, this shows that FACHB-351 is a robust producer of both biomass and phycocyanin (C-PC) under a wide array of culture conditions, especially with salinity of seawater or even higher (Jiang et al., 2019; Figure 1a). Although the accumulated biomass decreased with increasing salinity, FACHB-351 still grew at a salinity of 5.7%, which is almost twice that of seawater, namely 2.7–3.0% (Jiang et al., 2019). C-PC content exhibited a strong relationship with the salinity as well, unlike chlorophyll *a* (Chl *a*), whose content remained unchanged at various salinities. Since C-PC takes charge of light energy absorption and Chl *a* is responsible for energy transfer, the regulation of sink–source dynamics might depend on phycocyanin, rather than chlorophyll, when FACHB-351 encounters stresses.

Meanwhile, the phenotype of FACHB-351 has been maintained and exhibits potential for high-density cell accumulation by well-organized aggregation on solid surfaces of seawater-made plates (Figure 1b). This result indicates strong adaptive regulation in FACHB-351 to respond to saline environments. To fully understand the possible regulation pathways involved in salinity adaptation by FACHB-351, its genome sequence has been obtained.

2.2 Sequencing and assembly

The nucleotide sequence of the whole genome of *S. subsalsa*FACHB-351 was determined using Illumina HiSeq technologies, trimmed for quality, and then *de novo* assembled. After genome splicing and sequence correction, a genome assembly of 279 contigs and 5.2 Mbp was produced, and it comprises 4533 potential protein-coding genes, as well as 8 sets of rRNA genes, and 44 tRNA genes (Table 1).

S. subsalsa FACHB-351 contains 16 clustered, regularly interspaced short palindromic repeats (CRISPRs), which suggest a promising cellular defence against phages and plasmids and an enormous potential in the application of gene editing (Jinek et al., 2012; Cong et al, 2013). Table 1 also includes statistics on the contiguous sequences before assembly into the final scaffolds. The G+C% of the genome is 47.5%, which is close to the 47.7% recorded for *S. subsalsa* PCC 9445 (access no. GCA_000314005.1).

Translated amino acid sequences were compared with sequences in current mainstream databases, including NR, COG, GO and KEGG, using the BLAST program. Of the 4533 protein-coding sequences, 4289 (97%) are orthologous or had similarity to genes of known function or hypothetical genes. The genome of *S. subsalsa* FACHB-351 displays the highest overall synteny with the genomes of *Spirulina subsalsa* submitted by JGI (Shih et al., 2013), with 3,547 proteins as bidirectional best hits.

To determine biological roles of the assembled sequence, annotations in the COG, GO and KEGG databases were analysed. 'Signal transduction mechanisms' made up the third-highest percentage (10.5%) of annotated proteins in the COG database, following 'General function prediction only' (14.1%), and 'Function unknown' (12.7%). After 'Signal transduction mechanisms', 'Cell wall/membrane/envelope biogenesis' and

Posted on 10 Jan 2023 — The copyright holder is the author/funder. All rights reserved. No reuse without permission. — https://doi.org/10.22541/au.167393851.17288816/v1 — This a preprint and has not been peer reviewed. Data may be preliminary.

'Replication, recombination and repair' account for 7.3% and 6.5%, respectively (Figure 2). For GO annotation, terms with the 10 highest numbers of genes are starred in Figure 3, most of which relate to cell growth and adaptation to environments. Following from the results of COG and GO annotation, a set of genes code proteins that participate in cell energy metabolism and signal transduction. Recognising the elaborate pathways in KEGG, this database was used to further explore the genome features of *S. subsalsa* FACHB-351.

2.3 Comparative genomics

The cells of FACHB-351 have been morphologically and molecularly determined to be *Spirulina subsalsa* that is closely related to other *Spirulina* spp. and *Arthrospira platensis* (Figure S1). However, *S. subsalsa* FACHB-351 has few orthologues within the closely related species that have a fully sequenced genome, as shown in Table S1, which indicates specific proteins and their function in*S. subsalsa* FACHB-351.

Interestingly the whole 16S rRNA sequence of *S. subsalsa*FACHB-351 shows distances of 0.11–0.12 from *Pseudomonas* spp. in the NCBI's 16S database, without any closely related cyanobacteria (Table S2), which demonstrates unique gene functions of *S. subsalsa* FACHB-351, possibly analogous to heterotrophic bacteria rather than phototrophic cyanobacteria. By blasting protein sequences, 151 core orthologous clusters, only corresponding to 3.6% of total orthologous clusters, were found for listed *Pseudomonas* spp. and*S. subsalsa* FACHB-351, which further indicates strain-specific proteins in *S. subsalsa* FACHB-351. The function distribution of core orthologous clusters is shown in Figure 4a, with the greatest proportions (in decreasing order) assigned to 'Translation, ribosomal structure and biogenesis', 'Amino acid transport and metabolism' and 'Coenzyme transport and metabolism'.

To identify novel features of the *S. subsalsa* FACHB-351 genome, annotation with COG tools was also employed, which resulted in a distribution with 'Signal transduction mechanisms' as the third highest percentage in specific genes, following 'General function prediction only' and 'Function unknown' (Figure 4b).

These phylogenetic and orthologous cluster analyses manifest a particular position of *S. subsalsa* in the bacterial or cyanobacterial domain: (1) it has few orthologous clusters with morphologically and molecularly related strains of *Spirulina* spp. and *Arthrospira platensis* ; and (2) although *S. subsalsa*shows close distance to the *Pseudomonas* spp. bacteria based on 16S rRNA, they only share a small quantity of similar protein functions. Hence, the research on the genome sequence of *S. subsalsa*FACHB-351 would enrich the bacterial gene database and improve the diversity of information on evolution between bacteria and photosynthetic cyanobacteria.

2.4 Environmental information processing

2.4.1 Two-component systems

As a common system to sense and respond to environmental stresses in microbes, two-component systems (TCSs) canonically consist of two proteins, a histidine kinase (Hik) and a response regulator (Rre), which accounts for the most genes (259) in the genome of *S. subsalsa*FACHB-351. These sequences encode 19 putative genes for histidine kinases, including sensor histidine kinases, which are hybrid histidine kinases that contain not only a transmitter domain but also single or multiple receiver domain(s). There are also 90 putative genes for response regulators, such as phosphate or nitrogen regulation sensors.

Orthologous histidine kinases of *Synechocystis* , Hik2 (slr1147), Hik7 (SphS, sll0337), Hik8 (sasA, sll0750), Hik33 (sll0698) and Hik34 (slr1285), are conserved in almost all cyanobacterial genomes. The*S. subsalsa* FACHB-351 genome only contains orthologues of Hik8, a sensory histidine kinase necessary to sustain robust circadian timing in cyanobacteria (Takai et al., 2006). Considering the Hiks and its Rres, NblS–NblR and SasA–RpaA/RpaB could be the main TCS's to process information about light, circadian timing, or nutrient availability, and to regulate photosynthesis and metabolism.

2.4.2 cAMP and c-di-GMP signal transduction

Posted on 10 Jan 2023 — The copyright holder is the author/funder. All rights reserved. No reuse without permission. — https://doi.org/10.22541/au.167393851.17288816/v1 — This a preprint and has not been peer reviewed. Data may be preliminary.

cAMP is an important signalling molecule in cyanobacteria. The *S. subsalsa* FACHB-351 genome encodes 1 adenylate cyclase (K01768, E4.6.1.1) that produces cAMP from ATP, while cyanobacterial genomes usually encode less than 10 adenylate cyclases. Moreover, there is no enzyme encoded by *S. subsalsa* FACHB-351 to degrade cAMP, which reflected an incomplete synthesis and degradation pathway of cAMP. The cAMP signal transduction may be poorly developed in *S. subsalsa*FACHB-351 and may not be depended upon by *S. subsalsa* FACHB-351 to respond to external stimuli. There is no sequence to encode the enzyme that would synthesize another second messenger, c-di-GMP, while the related gene for cGMP-mediated signalling (GO:0019934) has been detected in the GO database.

2.4.3 Membrane transporters

Two subgroups of membrane were found in the FACHB-351 genome, namely ATP-binding cassette transporters (ABC transporters) and those in the bacterial secretion system. The substrates that can be transported with complete putative proteins in FACHB-351 may include phosphate, iron (III), urea, amino acids, and lipo-oligosaccharide. Besides that, the sequences for MsbA and HlyB are also annotated, and these proteins have been reported to have the ability to translocate hydrophobic drugs and lipids (Chang and Roth, 2006). Together with the genetically supposed presence of macrolide exporters (MacB and MsrA) in FACHB-351, its drug resistance could be hypothesized. This feature has been proved in several research studies, such as the survival of *S. subsalsa* at more than 300 μg/mL of neomycin (Zang et al., 2004). These genes in prokaryotic cyanobacterial FACHB-351 provide an efficient and easily implemented genetic tool to study the 'flip-flop' mechanism of drug movement across the lipid bilayer in disease treatment and development of novel therapeutics.

Regarding the common proteins submitted to the KEGG database for maintaining cellular ion balance, no gene was assigned to Na^+/H^+ antiporters or the Mrp-system. Within the GO database, 97, 12 and 9 genes have been found related to ion transport, sodium ion transport and solute:hydrogen antiporter activity, respectively. These results suggest that remarkable genes for transporters are located in the genome of FACHB-351, and some as-yet-undiscovered pathways for sodium to cross a membrane should exist in FACHB-351.

2.4.4 Chemotaxis

S. subsalsa FACHB-351 has complete pairwise Hiks–Rres in the chemotaxis family, which enables FACHB-351 to sense environmental signal by methyl-accepting chemotaxis protein (mcp, k03406) adjust motility and phytochrome for cell growth and development. Some chemotaxis-regulating*che* genes to methylate (*cheR*) or demethylate (*cheB*) mcp are also found in the genome of *S. subsalsa* FACHB-351, and they function as molecular memory for flagellar regulation, as confirmed in *E. coli* (Baker et al., 2006). The gene *cheC* is also detected as a phosphatase for the *cheY* response regulator (Rao et al., 2008). The sets of sequences related to molecular memory and a rapid turnover system of phosphorylation mean that FACHB-351 can avail itself of highly sensitive chemotactic responses to environmental stimuli.

Another gene cluster involved in regulation chemotaxis is *pil* , which encodes type IV pilins and is required for social (S)-motility (Li et al., 2003). The cells of an organism with S-motility can move together in coordination on a solid surface, which enables cells to enter a situation of extremely high cell density. Meanwhile, cells piled in organized layers atop one another also effectively contact each other for enhanced cell–cell communication and response to a signal (Dworkin et al., 1996). Behaviour characteristic of S-motility was still observed in *S. subsalsa* FACHB-351 on a solid surface composed of seawater (Figure 1b). These motility characteristics reflect a strong expression and regulation of genes in *S. subsalsa* FACHB-351 related to cell motility, alongside a good adaptation to suboptimal environments and potential for effective biomass generation.

2.5 Photosynthesis-related genes

Cyanobacteria, as phototrophs, convert light energy into chemical energy through photosynthetic complexes for cell growth and development. Thus it is critical for cyanobacteria to sense and respond to different light

Posted on 10 Jan 2023 — The copyright holder is the author/funder. All rights reserved. No reuse without permission. — https://doi.org/10.22541/au.167398511.17288816/v1 — This a preprint and has not been peer reviewed. Data may be preliminary.

environments, as well as to photosynthetically convert energy and fix carbon.

2.5.1 Phototaxis

Phototaxis is a kind of light-adaptation strategy that enables organisms to locate themselves in optimal light conditions and avoid UV irradiation or high-light exposure, which has been proved to be mediated by the *pix* gene cluster (*pixGHIJL*). For *S. subsalsa*FACHB-351 a total of 11 *pix* genes were detected to encode almost all *pix* proteins except *pixH* . The missing of *pixH*is consistent with the conclusion of Chen et al. (2020), who held that the absence of *pix* genes in major groups of aquatic cyanobacteria may be caused by changing light conditions under water. The protein*pixJ* belongs to TCS, but possessing different light-sensor domains and driving different cell responses to light information.

Together with *pix* genes, the light regulation of cyanobacterial growth usually depends on type IV pili, which can provide directional motive force (Bhaya et al., 2000; Wilde and Mullineaux, 2015). The presence of *pil* genes related to the biosynthesis of type IV pilins suggests the potential coordination from type IV pilus to phototactic movement.

2.5.2 Light absorption

In prokaryotic cyanobacteria, phycobilisomes — instead of the light-harvesting chlorophyll protein complexes in eukaryotes — function to harvest light and are composed mainly of phycobilin on phycocyanin (C-PC) and phycoerythrin (PE), together with the phycobilisome core of allophycocyanin (AP). The biosynthesis and regulation of phycocyanobilin is briefly shown in Figure 5, depicting the sequence (1) starting with L-glutamate for synthesizing protoporphyrin IX with the aid of HemALBCDEFY/G enzymes; (2) subsequently synthesizing Heme with enzyme HemH; (3) enzyme HO catalyzing the generation of biliverdin; (4) producing phycocyanobilin with the reductase pcyA; and (5) linking phycocyanobilin with apoprotein to form PC through enzyme CpcE/F or CpcST. Among the enzymes, known genes for heme oxygenase were not found in the genome of FACHB-351, while the fluorescence signal of C-PC has been observed on the excitation–emission fluorescence spectrum (Figure S2). Hence it is possible that a special enzyme with the same function as heme oxygenase exists in FACHB-351, which would need more evidence from proteomics to confirm. C-PC encoded and regulated by *cpc* genes can effectively convert red light, while PE encoded and regulated by *cpe* genes is better at absorbing green light. In the genome of FACHB-351 there are whole sets of sequences for synthesis and linking of C-PC and AP, while neither CpeA nor CpeB could be found. The absence of PE commonly occurs in more than 80% of non-marine cyanobacteria, since they usually have access to long-wavelength light and do not need to adjust pigment profiles for surviving in only deep-penetrating green and blue light, as marine cyanobacteria do in the ocean. This feature of phycobilisomes suggests that cultivation of FACHB-351 with green light would be inadvisable, as has been confirmed by our observation of 70% lower biomass production under green light than under red light (data not published).

Light energy harvested by C-PC will be transferred to chlorophyll*a* , where electrons are prepared for primary photochemical reaction and energize cell activities. The synthesis of chlorophyll*a* shares a precursor with phycocyanobilin, namely protoporphyrin IX, which suggests a potential competition between these two tetrapyrrole pigments. A vinyl reductase enzyme known to be essential, DVR, did not show up in the genome of FACHB-351, which suggests the presence of an alternative metabolic route and/or possibly variants of vinyl reductase genes that are beyond the scope of current knowledge. More research is necessary to explore whether there is a specific pathway synthesizing chlorophyll *a* , as with the reaction generating biliverdin in C-PC synthesis.

The decreased rate of C-PC synthesis was consistent with that of growth rate when the salinity was increasing in the media, while chlorophyll stayed constant (see Figure 1a). The enzymes involved in synthesizing C-PC are possibly more sensitive to ambient change than those related to chlorophyll; moreover, this sensitivity could assist FACHB-351 to maintain the source–sink balance and further exhibit robust growth under stresses such as those due to high salinity.

2.5.3 Photosynthesis

Posted on 10 Jan 2023 — The copyright holder is the author/funder. All rights reserved. No reuse without permission. — https://doi.org/10.22541/au.167393851.17288816/v1 — This a preprint and has not been peer reviewed. Data may be preliminary.

Almost all photosynthesis genes were identified in *S. subsalsa*FACHB-351, including photosystem I and II (PS I and II), cytochromeb_6-f complex and ATP synthase (Figure 6). For photosystem I, all known prokaryotic genes except *psaX* were detected. The function of PsaX in PS I is to form a trimeric reaction centre with PsaA and PsaB, but its presence has not been confirmed in thermophilic or low-light adapted species (Krauf, 2008; Nelson, 2011). The absence of PsaX in mesophilic *S. subsalsa* FACHB-351, as for mesophilic *Synechocystis* sp. PCC6803, provides another example supporting the hypothesis of PsaX functioning in improving PS I thermostability (Krauf, 2008). For photosystem II, the complete sequences for reaction centre protein, chlorophyll apoprotein and oxygen-evolving enhancer protein were detected, while several low molecular mass subunits (*psbLSWY*) were missing in the genome of FACHB-351. These subunits work as regulators of electron transfer in PS II, and some of them can collaborate or work in alternation, such as*psbJ* to *psbL* (Ohad et al., 2004). The presence of*psbJ* might be the reason for the effective photosynthesis by*S. subsalsa* FACHB-351 without *psbLSWY* genes. *PetE*and *PetJ* encode electron-carriers plastocyanin and cytc_6, respectively, and plastocyanin works with copper-dependence while cyt c_6 can replace plastocyanin in the absence of copper (Kerfeld and Krogmann, 1998). The two genes were found in the genome of FACHB-351, which further indicates a strong regulation of electron-transfer efficiency when cultures encounter copper limitation.

2.6 Carbohydrates

In the KEGG database the metabolism group was located by the highest gene number from the *S. subsalsa* FACHB-351 sequence, and carbohydrate metabolism exhibited the highest gene number among the subgroups (Figure S3). Hence plentiful carbohydrate-controlled enzymes were found in the Carbohydrate-Active Enzymes database (CAZy), as shown in Figure 7, which indicates the presence of luxury carbohydrates with diverse structures, as well as huge possibilities for the medication. The most enriched CAZy class is glycosyltransferases (GTs), for which the top five enriched families were GT41, GT4, GT5, GT2 and GT27. All of them relate to monosaccharide modification of protein residues, synthesis of glycogen, sucrose, cellulose and chitin.

In cyanobacteria, the carbohydrates metabolised from oligosaccharides and polysaccharides function as cell wall components or stress metabolites. Examples of oligosaccharides are sucrose, trehalose and glucosylglycerol (GG) for salt tolerance. Three putative trehalose synthesis pathways have been detected in the genome of *S. subsalsa* FACHB-351: (1) *de novo* synthesis starting from UDP-glucose or glucose-6-phosphate successively through trehalose phosphate synthase (TPS) and trehalose-phosphate phosphatase (TPP); (2) converting glycogen/starch by glycogen debranching enzyme, maltooligosyl trehalose synthase (Mts) and maltooligosyl trehalose hydrolase (Mth); (3) a single transglycosylation reaction catalyzed by trehalose synthase (TreS). Among them the first route is the most common one, while the other two are less prominent (Wolf, et al., 2003). The presence of genes encoding enzymes for every route of trehalose synthesis suggests that*S. subsalsa* FACHB-351 possesses a powerful system and further indicates a stronger resistance to stress (Figure 7b).

GG is a pervasive osmoprotectant compound and synthesized only from ADP-glucose under the catalyzation of GG-phosphate synthase and GG-phosphate phosphatase (Ball et al., 2015). Two genes for GG-phosphate synthase exist in *S. subsalsa* FACHB-351, while none are present for GG-phosphate phosphatase. However, strong adaptation to salinity leads us to think that there might be some unknown type of enzyme with the same function as GG-phosphate phosphatase, or that there is strong sucrose accumulation. A large amount of sucrose was measured after osmotic shock to mutant *Synechocystis* PCC6803 that ineffectively produces GG, which was interpreted as a compensation for lack of GG in the osmotic balance process (Miao et al. 2003).

For sucrose synthesis, cyanobacteria commonly employ fructose-6-phosphate and UDP-glucose as substrates and a two-step pathway sequentially involving sucrose-phosphate synthase (SPS) and sucrose-phosphate phosphatase (SPP) (Figure 7c). Many genes encoding SPS were sequenced in the genome of *S. subsalsa* FACHB-351, while none for SPP were detected. Beside the two aforementioned enzymes for sucrose yield, Martínez-Noël et al. (2013) demonstrated another possible route in the cyanobacterium *Synechococcus elon-*

Posted on 10 Jan 2023 — The copyright holder is the author/funder. All rights reserved. No reuse without permission. — https://doi.org/10.22541/au.167393851.17288816/v1 — This a preprint and has not been peer reviewed. Data may be preliminary.

gatus PCC 7942 that could be conducted by only one bidomainal SPS with both SPS and SPP activity (Figure 7c). Moreover, this one-enzyme route was proven to be more effective than the two-step pathway (Martínez-Noël et al., 2013). This type of bidomainal enzyme might exist in *S. subsalsa* FACHB-351 as well, since sucrose has been detected in it. In this view, the absence of sequence for a putative SPP, together with sucrose detected in*S. subsalsa* FACHB-351, indicates a possible swift synthesis of sucrose and then an efficient route compensating for lack of GG to achieve osmotic balance.

2.7 Virulence factors of pathogenic bacteria

To evaluate the feasibility of *S. subsalsa* FACHB-351 as a potential feedstock for food, animal feed and pharmaceutical use, analysis of the virulence factors of pathogenic bacteria was performed with Virulence Factor Database (VFDB). Although in total 416 proteins are found to be related to virulence, no bacterial toxins showed up; meanwhile, their E-value was below 10^{-6}, which suggests a low risk for *S. subsalsa* FACHB-351 to be toxic to humans or animals. Most of these blasted proteins are involved in the processes of secretion, cell attachment and motility that facilitate cell communication and protect themselves from suboptimal ambient conditions. For example, a halotolerant gene for hemolysin B (*hlyB*) identified in *S. subsalsa* FACHB-351 has been proved to be a gene governing cyanobacteria's response to osmotic stress. Sakiyama et al. (2011) and Tabatabai et al. (2017) found that hemolysin overproduction helped *Synechocystis* sp. PCC 6803 and*Fremyella diplosiphon* thrive at high salinity, since hemolysin-like proteins localized in the S-layer of the organism participate in maintaining cell shape under stressful conditions (Uchiyama et al., 2015). A possible effective regulation of this gene expression might confer halotolerance of *S. subsalsa* FACHB-351 in salinity of 2.7–5.7% — equal to or even higher than that of seawater (Figure 1).

3. Discussion

A cyanobacterium, *S. subsalsa* FACHB-351, with strong adaptation to salinity and high PC yields has been reported and exhibits promising prospects in the field of PC production, together with extending the knowledge of microbial responses to hypersaline environments. The paucity of orthologous clusters between *S. subsalsa* FACHB-351 and its morphologically and molecularly related species suggest a particular position of *S. subsalsa* in the bacterial or cyanobacterial domains (Figure 4a). Hence our research on the genome sequence of*S. subsalsa* FACHB-351 will enrich the bacterial gene database and improve the diversity of knowledge about evolutionary development between bacteria and photosynthetic cyanobacteria, and will also permit comprehensive approaches using systems biology to develop freshwater-free cultivation strategies for commercial production of phycocyanin from *S. subsalsa* .

Among all the annotated proteins from the assembled sequence we focused on some related to environmental information processing, photosynthesis, and carbohydrate metabolites. The genes employed to synthesize proteins related to environmental adaptation are the most abundant, expressed as signal transduction (Figure 2 and 3). However, the FACHB-351 sequence was not assigned to some known signal messengers and membrane transporters in KEGG, while the sequence was blasted to corresponding proteins in GO. Some known genes of essential enzymes that participate in primary cell activities, such as HO and DVR for the synthesis of PC and chlorophyll, were likewise missing in the genome of*S. subsalsa* FACHB-351 during the process of constructing pathways with KEGG. This situation suggests the presence of possible special and as yet undiscovered enzymes in FACHB-351 and inspires further research employing proteomics and transcriptomics. More exploration of FACHB-351 under suboptimal culture conditions might enrich the knowledge of cyanobacterial metabolism.

With overall consideration of gene distribution and growth observation, cellular defence mechanisms of FACHB-351 to saline stress include, but are not limited to: (1) A well-developed signal transduction system, with the most abundant gene distribution; (2) Maintaining a source–sink balance, depending on the synthesis and behaviour of PC; (3) Keeping a cellular ion balance through ion transporters and disaccharide osmoprotectants; and (4) Forming layered cell aggregation by S-motility that facilitates cell–cell communication and response.

Analysis using VFDB showed no genes encoding bacterial toxins in the genome of FACHB-351, which

Posted on 10 Jan 2023 — The copyright holder is the author/funder. All rights reserved. No reuse without permission. — https://doi.org/10.22541/au.167398511.17288816/v1 — This a preprint and has not been peer reviewed. Data may be preliminary.

provides preliminary proof of FACHB-351's safety as a potential feedstock for phycocyanin.

4. Materials and methods

4.1 Microalgal strain and culture conditions

The cyanobacterium *S. subsalsa* used in this study was purchased from the Freshwater Algae Culture Collection of the Institute of Hydrobiology (FACHB-351). *Spirulina* medium (SP) was recommended by FACHB, and it consisted of (per litre of distilled water): 13.61 g $NaHCO_3$, 4.03 g Na_2CO_3, 0.5 g K_2HPO_4, 2.50 g $NaNO_3$, 1.00 g K_2SO_4, 1.00 g NaCl, 0.20 g $MgSO_4$*$7H_2O$, 0.04 g $CaCl_2$*$2H_2O$, 0.04 g $FeSO_4$*$7H_2O$ and 1 mL of A5. A5 trace metal solution consists of (per litre of distilled water): 2.86 g H_3BO_3, 1.86 g $MnCl_2$*$4H_2O$, 0.22 g $ZnSO_4$*$7H_2O$, 0.39 g Na_2MoO_4*$2H_2O$, 0.08 g $CuSO_4$*$5H_2O$ and $Co(NO_3)_2$*$6H_2O$.

A series of salinities were obtained by adding different concentrations of NaCl (20 to 50 g/L, at intervals of 10 g/L) into SP to observe the response of FACHB-351 to salt stress. Flasks of 250 mL capacity each containing 150 mL of SP were inoculated with the same amount of inoculum to make an initial biomass concentration of around 0.1 g/L. All flasks were kept at a temperature of 25 +- 2 degC and a light intensity around 60 μmol/m^2/s for 24 h. All of the batch experiments were conducted in triplicate.

4.2 Culture on solid surface

Solid seawater-based plates were prepared by adding 1% agar into seawater supplemented with monosodium glutamate residue, as previously reported (Jiang et al., 2019). Cyanobacterial inoculum with biomass concentration of 0.5 g/L was obtained by centrifuging exponential cells of FACHB-351 in SP, washing the cells with deionized water and resuspending in deionized water. Spots were achieved by dropping 50 μL of the abovementioned inoculum at the centre of plates and incubating for around 4 h until no observable liquid remained. Plates were then grown under the same conditions as the flask cultures. The observation of spots was carried out under a microscope (CX31, Olympus, Japan).

4.3 Biomass, phycocyanin and chlorophyll measurement

After 5-day cultivation, all culture in a 250-mL flask was centrifuged at 4000 rpm for 5 min and washed 3 times with deionized water to obtain a cyanobacterial pellet. The pellet was completely transferred into a weighed ziplock bag for drying in a lyophilizer (EYELA FDU-1200, Tokyo Rikakikai, Japan). Then the ziplock bag containing dried cyanobacterial biomass was weighed to determine the dry cell mass (DCM).

For chlorophyll *a* (Chl *a*) determination, around 2 mg of dry biomass was weighed and combined with 4 mL of methanol (99%). After vortex-mixing, the mixture of cells and methanol was stored in the dark for 24 h at 45 °C. The methanol extract was obtained through centrifugation at 4000 rpm for 10 min, and then the absorbances at wavelengths of 665.2 and 652.4 nm were found with the use of a UV–vis spectrophotometer (UV-2450, Shimadzu, Japan). The Chl *a* content was calculated as:

$$\text{Chl } a \text{ (\% of DCM)} = \frac{(16.72A_{665.2} - 9.16A_{652.4}) \times 4}{m_1} \times \frac{100}{1000} \quad (1)$$

where m_1 (mg) stands for the mass of dry biomass used for Chl *a* measurement.

For C-PC extraction, a fixed amount of the dried biomass (2 mg DCM) was suspended in 4 mL of phosphate-buffered saline (PBS) and the solution was maintained at -20 degC for 8 h. After that, the frozen sample was thawed in a sonication bath and the cell debris was then removed by centrifugation at 4000 rpm for 5 min, and the supernatant (of blue colour) was collected. The absorbance of crude extract was measured on a UV–vis spectrophotometer at wavelengths of 620 and 652 nm, and the C-PC content was calculated according to Eq. (2) (Bennett and Bogorad, 1973):

$$\text{C-PC (\% of DCM)} = \frac{A_{620} - 0.474\ A_{652}}{5.34} \times \frac{4}{m_2} \times 100 \quad (2)$$

where m_2 (mg) stands for the mass of dry biomass used for C-PC measurement.

4.4 DNA sequencing and genome assembly

Posted on 10 Jan 2023 — The copyright holder is the author/funder. All rights reserved. No reuse without permission. — https://doi.org/10.22541/au.167393851.17288816/v1 — This a preprint and has not been peer reviewed. Data may be preliminary.

Axenic cells of *Spirulina subsalsa* FACHB-351 cultured in SP were pelleted by centrifugation and washed twice with PBS. With the pelleted cells, genome DNA was extracted, sequenced, and assembled by Sangon Biotech (Shanghai) Co. (Shanghai, China). Briefly, genomic DNA was isolated using magnetic nanoparticle technologies and sequenced by Illumina HiSeq. Raw data were quality checked and assembled by FastQC (v. 0.11.2), Trimmomatic (v. 0.36) (Bolger et al., 2014), SPAdes (v. 3.5.0) (Bankevich et al., 2012), GapFiller (v. 1.11) (Boetzer and Pirovano, 2012), and PrInSeS-G (v. 1.0.0) (Massouras et al., 2010). This Whole Genome Shotgun project has been deposited with NCBI/GenBank under the accession code *JAIHOM000000000* .

4.5 Genome analysis and annotation

Protein-coding genes, tRNA and rRNA, were identified using the Prokka program (v. 1.10) (Seemann, 2014), whereas CRISPRs were identified using CRT (v. 1.2) (Bland et al., 2007). For functional annotation, the NCBI NR database, NCBI COG database and GO database were searched to assign the predicted protein sequences based on sequence similarities. The KEGG database was used for pathway reconstruction. The virulence potential was analyzed with the Virulence Factor Database (VFDB). Carbohydrate-active enzymes in proteins were predicted using HMMER3 (v. 3.1b1) (Eddy, 2009) and the CAZy database with a filter of E-value lower than 10^{-5} (www.cazy.org).

Phylogenetic relationships among same-family cyanobacteria and 16S rRNA-close species were manually curated using Blast, muscle and FastTree software. Orthologous genes were analyzed with the PGAP program (v. 1.2.1) under the constraint for identity higher than 95 (Zhao et al., 2012). The genomic information and gene sequence of related strains were obtained from the NCBI RefSeq database (Table S1 and S2).

Declaration

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Availability of data and materials

The datasets generated for this study can be found in National Center for Biotechnology Information (NCBI) under the accession numbers: under the accession code*JAIHOM000000000* .

Competing interests

The authors declare that they have no conflict of interest.

Funding

This work was supported by the China Postdoctoral Science Foundation [grant number 2020M682186] and Key Technology Research and Development Program of Shandong (International Cooperation) [grant number 2019GHZ030].

Authors' contributions

Liqun Jiang: Conceptualization, Investigation, Writing- Original draft preparation. Hangzhou Xu: Investigation. Haiyan Pei: Supervision, Writing- Reviewing and Editing.

Acknowledgments

We would like to thank Dr. David Verrelli (Division One Academic and Language Services) for kindly editing the grammar and language.

References

Posted on 10 Jan 2023 — The copyright holder is the author/funder. All rights reserved. No reuse without permission. — https://doi.org/10.22541/au.167398511.17288816/v1 — This a preprint and has not been peer reviewed. Data may be preliminary.

1. Ball S, Colleoni C, Arias MC. The transition from glycogen to starch metabolism in cyanobacteria and eukaryotes. Starch, 2015, 93–158. http://dx.doi.org/10.1007/978-4-431-55495-0_4
2. Baker, MD, Wolanin PM, Stock JB. Systems biology of bacterial chemotaxis. Curr Opin Microbiol, 2006, 9:187–192. https://doi.org/10.1016/j.mib.2006.02.007
3. Bankevich A, Nurk S, Antipov D, Gurevich AA, Dvorkin M, KulikovAS, Lesin VM, Nikolenko SI, Pham S, Prjibelski AD, Pyshkin AV, Sirotkin AV, Vyahhi N, Tesler G, Alekseyev MA, Pevzner PA. SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. J Comput Biol, 2012, 19:455–477. https://doi.org/10.1089/cmb.2012.0021
4. Bhaya D, Bianco NR, Bryant D, Grossman A. Type IV pilus biogenesis and motility in the cyanobacterium *Synechocystis* sp. PCC6803. Mol Microbiol, 2000, 37:941–951. https://doi.org/10.1046/j.1365-2958.2000.02068.x
5. Bland C, Ramsey TL, Sabree F, Lowe M, Brown K, Kyrpides NC, Hugenholtz P. CRISPR recognition tool (CRT): a tool for automatic detection of clustered regularly interspaced palindromic repeats. BMC Bioinformatics, 2007, 8:209. https://doi.org/10.1186/1471-2105-8-209
6. Boetzer M, Pirovano W. Toward almost closed genomes with GapFiller. Genome Biol, 2012, 13:R56. https://doi.org/10.1186/gb-2012-13-6-r56
7. Bolger AM, Lohse M, Usadel B. Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics, 2014, 30:2114–2120. https://doi.org/10.1093/bioinformatics/btu170
8. Chang G, Roth CB. Structure of MsbA from *E. coli* : a homolog of the multidrug resistance ATP binding cassette (ABC) transporters. Science, 2001, 293:1793–1800. https://doi.org/10.1126/science.293.5536.1793
9. Chen MY, Teng WK, Zhao L, Hu CX, Zhou YK, Han BP, Song LR, Shu WS. Comparative genomics reveals insights into cyanobacterial evolution and habitat adaptation. ISME J, 2021, 15:211–227. https://doi.org/10.1038/s41396-020-00775-z
10. Cong L, Ran F, Cox D, Lin S, Barretto R, Habib N, Hsu P, Wu X, Jiang W, Marraffini L, Zhang F. Multiplex genome engineering using CRISPR/Cas systems. Science, 2013, 339: 819–823. https://doi.org/10.1126/science.1231143
11. Cui J, Sun T, Chen L, Zhang W. Engineering salt tolerance of photosynthetic cyanobacteria for seawater utilization. Biotechnol Adv, 2020, 43:107578. https://doi.org/10.1016/j.biotechadv.2020.107578
12. Dworkin M. Recent advances in the social and developmental biology of the myxobacteria. Microbiol Rev, 1996, 60:70–102. https://doi.org/10.1128/mr.60.1.70-102.1996
13. Eddy SR. A new generation of homology search tools based on probabilistic inference. Genome Inform, 2009, 23:205–211. https://doi.org/10.1142/9781848165632_0019
14. Gorelova V, Bastien O, De Clerck O, Lespinats S, Rebeille F, Van Der Straeten D. Evolution of folate biosynthesis and metabolism across algae and land plant lineages. Sci Rep, 2019 9:1–16. https://doi.org/10.1038/s41598-019-42146-5
15. Jinek M, Chylinski K, Fonfara I, Hauer M, Doudna J, Charpentier E. A programmable dual-RNA–guided DNA endonuclease in adaptive bacterial immunity. Science, 2012, 337:816–821. https://doi.org/10.1126/science.1225829
16. Kerfeld CA, Krogmann DW. Photosynthetic cytochromes *c* in cyanobacteria, algae, and plants. Annu Rev Plant Physiol Plant Mol Biol, 1998, 49:397–425. https://doi.org/10.1146/annurev.arplant.49.1.397
17. Krauf N. Structure and function of cyanobacterial photosystem I. Photosynthetic protein complexes, a structural approach, edited by Petra Fromme, 2008, 23–51. https://doi.org/10.1002/9783527623464.ch2
18. Li Y, Sun H, Ma X, Lu A, Lux R, Zusman D, Shi W. Extracellular polysaccharides mediate pilus retraction during social motility of*Myxococcus xanthus* . P Natl Acad Sci USA, 2003, 100:5443–5448. https://doi.org/10.1073/pnas.0836639100
19. Martinez-Noel GMA, Cumino AC, Kolman MA, Salerno GL. First evidence of sucrose biosynthesis by single cyanobacterial bimodular proteins. FEBS Lett, 2013 587:1669–1674. https://doi.org/10.1016/j.febslet.2013.04.012
20. Massouras A, Hens K, Gubelmann C, Uplekar S, Decouttere F, Rougemont J, Cole ST, Deplancke B.

Posted on 10 Jan 2023 — The copyright holder is the author/funder. All rights reserved. No reuse without permission. — https://doi.org/10.22541/au.167398511.17288816/v1 — This a preprint and has not been peer reviewed. Data may be preliminary.

Primer-initiated sequence synthesis to detect and assemble structural variants. Nat Methods, 2010, 7:485–486. https://doi.org/10.1038/nmeth.f.308

21. Miao X, Wu Q, Wu G, Zhao N. Sucrose accumulation in salt-stressed cells of *agp* gene deletion-mutant in cyanobacterium*Synechocystis* sp. PCC 6803. FEMS Microbiol Lett, 2003, 218:71–77. https://doi.org/10.1111/j.1574-6968.2003.tb11500.x
22. Nelson N. Photosystems and global effects of oxygenic photosynthesis. BBA–Bioenergetics, 2011, 1807: 856–863. https://doi.org/10.1016/j.bbabio.2010.10.011
23. Ohad I, Bosco CD, Herrmann RG, Meurer J. Photosystem II proteins PsbL and PsbJ regulate electron flow to the plastoquinone pool. Biochemistry, 2004, 43:2297–2308. https://doi.org/10.1021/bi0348260
24. Rao CV, Glekas, GD, Ordal GW. The three adaptation systems of*Bacillus subtilis* chemotaxis. Trends Microbiol 2008, 16:480–487. https://doi.org/10.1016/j.tim.2008.07.003
25. Sakiyama H, Araie I, Suzuki Y. Functions of a hemolysin-like protein in the cyanobacterium *Synechocystis* sp. PCC 6803. Arch Microbiol, 2011, 193:565–571. https://doi.org/10.1007/s00203-011-0700-2
26. Seemann T. Prokka: rapid prokaryotic genome annotation. Bioinformatics, 2014, 30:2068–2069. https://doi.org/10.1093/bioinformatics/btu153
27. Shih PM, Wu D, Latifi A, Axen SD, Fewer DP, Talla E, Calteau A, Cai F, De Marsac NT, Rippka R, Herdman M, Sivonen K, Coursin T, Laurent T, Goodwin L, Nolan M, Davenport KW, Han CS, Rubin EM, Eisen JA, Woyke T, Gugger M, Kerfeld CA. Improving the coverage of the cyanobacterial phylum using diversity-driven genome sequencing. P Natl Acad Sci USA, 2013, 110:1053–1058. https://doi.org/10.1073/pnas.1217107110
28. Tabatabai B, Arumanayagam ACS, Enitan O, Mani A, Natarajan SS, Sitther V. Overexpression of *hlyB* and *mdh* genes confers halotolerance in *Fremyella diplosiphon* , a freshwater cyanobacterium. Enzyme Microb Tech, 2017 103:12–17. https://doi.org/10.1016/j.enzmictec.2017.04.009
29. Takai N, Nakajima M, Oyama T, Kito R, Sugita C, Sugita M, Kondo T, Iwasaki H. A KaiC-associating SasA–RpaA two-component regulatory system as a major circadian timing mediator in cyanobacteria. P Natl Acad Sci USA, 2006, 103:12109–12114. https://doi.org/10.1073/pnas.0602955103
30. Uchiyama J, Kanesaki Y, Iwata N, Asakura R, Funamizu K, Tasaki R, Agatsuma M, Tahara H, Matsuhashi A, Yoshikawa H, Ogawa S. Genomic analysis of parallel-evolved cyanobacterium *Synechocystis* sp. PCC 6803 under acid stress. Photosynth Res, 2015, 125:243–254. https://doi.org/10.1007/s11120-015-0111-3
31. Wilde A, Mullineaux CW. Motility in cyanobacteria: polysaccharide tracks and Type IV pilus motors. Mol Microbiol, 2015, 98:998–1001. https://doi.org/10.1111/mmi.13242
32. Wolf A, Kramer R, Morbach S. Three pathways for trehalose metabolism in *Corynebacterium glutamicum* ATCC13032 and their significance in response to osmotic stress. Mol Microbiol, 2003, 49:1119–1134. https://doi.org/10.1046/j.1365-2958.2003.03625.x
33. Zang X, Zhang X, Wang G, Liu J, Zhang X, Liu B. Comparative study on sensitivities of *Arthrospira platensis* and *Spirulina subsulsa* to seven antibiotics. Acta Oceanol Sin, 2004, 26(3):82–92. https://doi.org/10.3321/j.issn:0253-4193.2004.03.010
34. Zhao Y, Wu J, Yang J, Sun S, Xiao J, Yu J. PGAP: pan-genomes analysis pipeline. Bioinformatics, 2012, 28:416–418. https://doi.org/10.1093/bioinformatics/btr655

Table 1. Genome statistics of *S. subsalsa* FACHB-351.

Characteristic	Value
Assembly length	5,264,922 bp
G+C content	47.47 %
Protein coding genes	4,533
Total coding gene length	4,389,015 bp
Coding ratio	83.92%
Non-coding RNA	
tRNA	44
rRNA	8

Posted on 10 Jan 2023 — The copyright holder is the author/funder. All rights reserved. No reuse without permission. — https://doi.org/10.22541/au.167393851.17288816/v1 — This a preprint and has not been peer reviewed. Data may be preliminary.

Characteristic	Value
Repeat regions	
Repeat region count	1,859
Total repeat region	181,438 bp
Repeat ratio	3.47 %
Annotated in at least one database	4,247
Annotated in COG	2,919
Annotated in GO	2,620
Annotated in KEGG	1,670

Figure captions

Figure 1. Growth of *S. subsalsa* FACHB-351 in saline environment. (a) FACHB-351 production of biomass, C-PC and Chl *a* in SP supplemented with extra NaCl. (b) Microscopic image of FACHB-351 on solid surface of seawater-based medium.

Figure 2. Assembled genome annotation of *S. subsalsa* FACHB-351 in the COG database.

Figure 3. Assembled genome annotation of *S. subsalsa* FACHB-351 in the GO database. Genes ranking within the top 10 in each ontology are included in this figure. Red stars identify the genes in the top 10 overall (by number).

Figure 4. Gene annotation of orthologues in the COG database: (a) core and (b) group-specific clusters between *S. subsalsa* FACHB-351 and closely related strains (by 16S rRNA).

Figure 5. Related pathways to light absorption of *S. subsalsa*FACHB-351. (a) The biosynthesis of phycocyanin and chlorophyll *a*starting with l-glutamate, where the genes for enzymes shown in red were not found in the genome of *S. subsalsa* FACHB-351. (b) The genes that code enzymes shaded in green were found in the genome of*S. subsalsa* FACHB-351.

Figure 6. Pathways related to photosynthesis by *S. subsalsa*FACHB-351, where the genes that code enzymes shaded in green were found in the genome of *S. subsalsa* FACHB-351. Red arrows stand for chemical reactions, red dashed arrows represent electron transport, and blue arrows indicate transmembrane movement of hydrogen ions.

Figure 7. Pathways related to carbohydrate metabolism by *S. subsalsa* FACHB-351: (a) assembled genome annotation of *S. subsalsa* FACHB-351 in the CAZy database; (b and c) putative pathways for synthesizing trehalose and sucrose, where the genes for enzymes shown in red were not found in the genome of *S. subsalsa*FACHB-351.

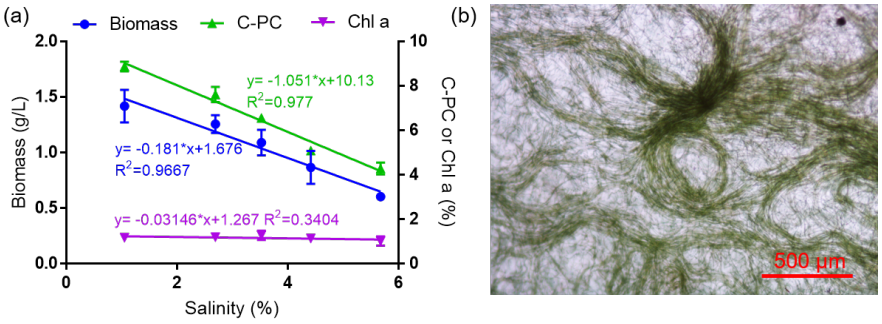

Figure 1

Posted on 10 Jan 2023 — The copyright holder is the author/funder. All rights reserved. No reuse without permission. — https://doi.org/10.22541/au.167393851.17288816/v1 — This a preprint and has not been peer reviewed. Data may be preliminary.

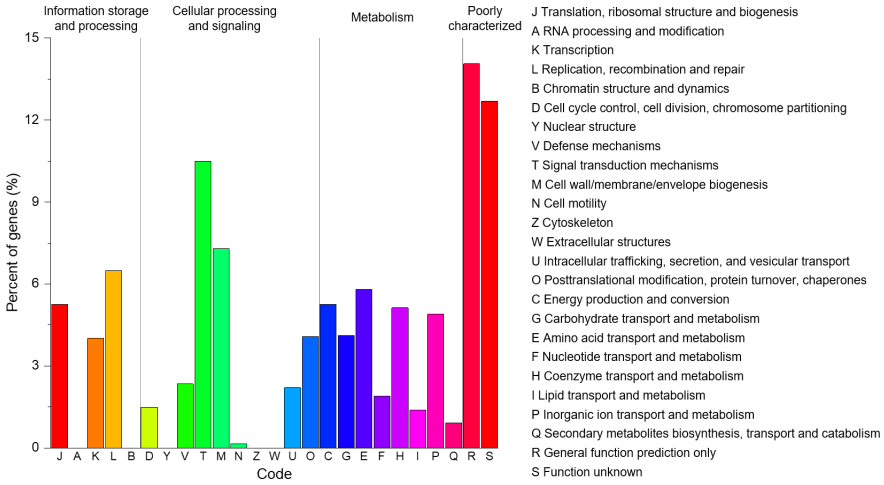

Figure 2

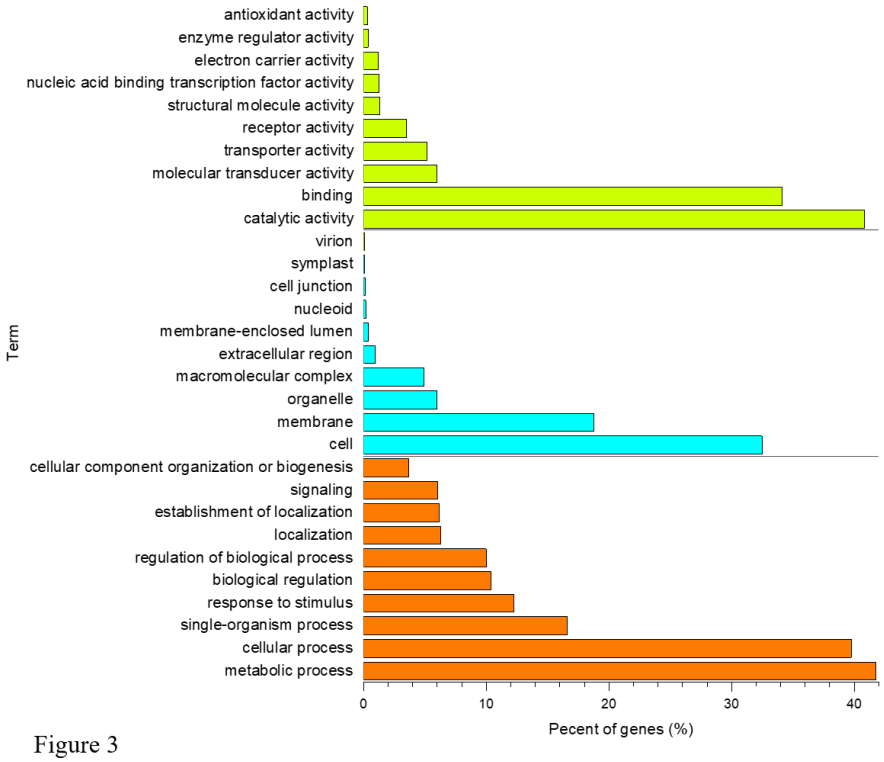

Figure 3

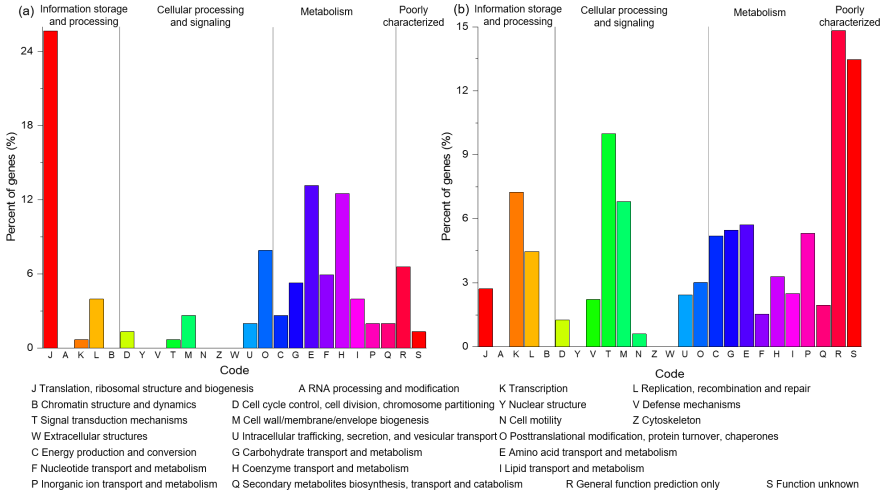

Figure 4

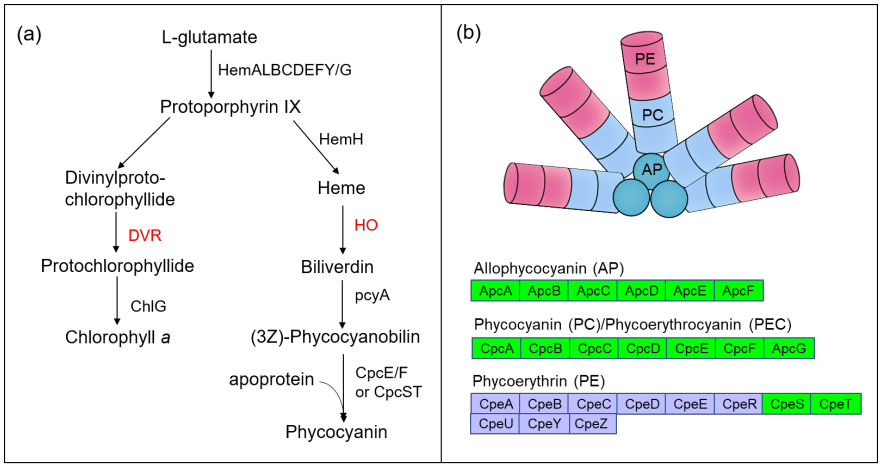

Figure 5

Posted on 10 Jan 2023 — The copyright holder is the author/funder. All rights reserved. No reuse without permission. — https://doi.org/10.22541/au.167393851.17288816/v1 — This a preprint and has not been peer reviewed. Data may be preliminary.

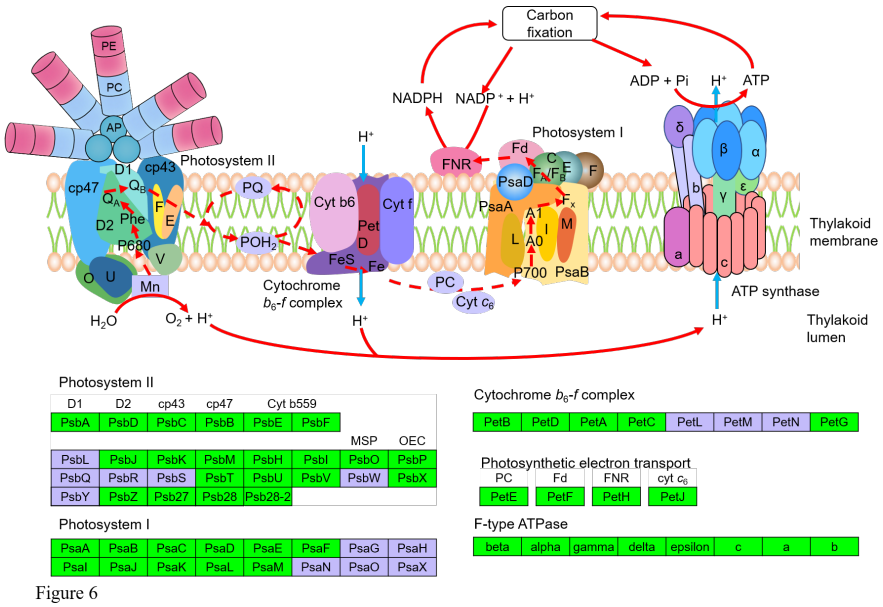

Figure 6

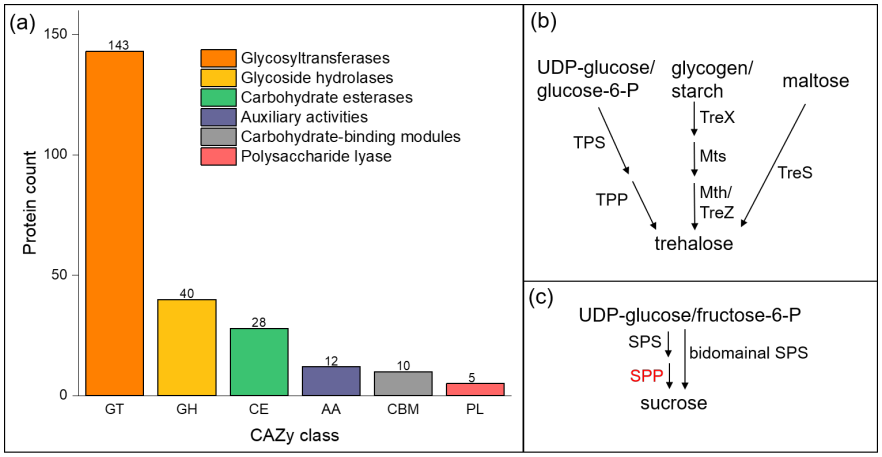

Figure 7

Posted on 10 Jan 2023 — The copyright holder is the author/funder. All rights reserved. No reuse without permission. — https://doi.org/10.22541/au.167398511.17288816/v1 — This a preprint and has not been peer reviewed. Data may be preliminary.