

EXTENDED GENOME REPORT

Open Access



Complete genome sequence and genomic characterization of *Microcystis panniformis* FACHB 1757 by third-generation sequencing

Jun-Yi Zhang^{1,2†}, Rui Guan^{1†}, Hu-Jun Zhang², Hua Li³, Peng Xiao⁴, Gong-Liang Yu³, Lei Du⁵, De-Min Cao⁵, Bing-Chuan Zhu², Ren-Hui Li³ and Zu-Hong Lu^{1,6*}

Abstract

The cyanobacterial genus *Microcystis* is well known as the main group that forms harmful blooms in water. A strain of *Microcystis*, *M. panniformis* FACHB1757, was isolated from Meiliang Bay of Lake Taihu in August 2011. The whole genome was sequenced using PacBio RS II sequencer with 48-fold coverage. The complete genome sequence with no gaps contained a 5,686,839 bp chromosome and a 38,683 bp plasmid, which coded for 6,519 and 49 proteins, respectively. Comparison with strains of *M. aeruginosa* and some other water bloom-forming cyanobacterial species revealed large-scale structure rearrangement and length variation at the genome level along with 36 genomic islands annotated genome-wide, which demonstrates high plasticity of the *M. panniformis* FACHB1757 genome and reveals that *Microcystis* has a flexible genome evolution.

Keywords: *Microcystis panniformis* FACHB1757, *Microcystis*, Lake Taihu, Water bloom, Third-generation sequencing, Comparative genomics

Introduction

The massive development of bloom-forming cyanobacteria is causing problems in eutrophic water bodies worldwide. Among the cyanobacteria, *Microcystis* is perhaps the most notorious. Many *Microcystis* species have been reported to be able to produce microcystins [1–4], which threaten many aquatic ecosystems and cause serious and occasionally fatal human liver, digestive, neurological, and skin diseases [5–7].

Microcystis is a genus of unicellular colony-forming cyanobacteria whose taxonomy is still unclear [8]. Although morphological criteria have been proposed to distinguish *Microcystis* species from field samples, such criteria have long been questioned for use in species identification within the genus [9]. Several studies

attempted to reconcile molecular and morphological taxonomy in *Microcystis* [9–14], and a morphology-based taxonomic system has been dominantly used. *Microcystis panniformis* was first reported in 2002 and was morphologically described as having flattened, irregular, monolayer colonies with small holes inside and later disintegrated into small pieces [15]. Since the *M. panniformis* strain SPC 702 was successfully isolated from Lago das Garças, São Paulo in 1999, studies addressing different aspects of this species have been performed [16–25]. In China, *M. panniformis* was reported as a newly recorded species in 2012 [26], and one strain (FACHB1757) was isolated from Lake Taihu. *Microcystis panniformis* was originally thought to only be distributed in tropical regions, but we showed that this species has invaded the subtropical regions with a monsoon climate [26]. Global expansion of harmful cyanobacteria has been thought to be linked to climate changes, particularly increasing amounts of atmospheric CO₂ and surface temperature, which may promote *Microcystis* growth and enhance the potential for bloom

* Correspondence: zhlu@seu.edu.cn

†Equal contributors

¹State Key Laboratory of Bioelectronics, School of Biological Sciences and Medical Engineering, Southeast University, Nanjing 210096, China

⁶Department of Biomedical Engineering, College of Engineering, Peking University, Beijing 100871, China

Full list of author information is available at the end of the article

occurrence [27–29]. Therefore, a deeper understanding of the ecology and physiology of *M. panniformis* FACHB1757 by obtaining a robust genome reference may provide insight into the expansion and invasion mechanisms of *Microcystis*.

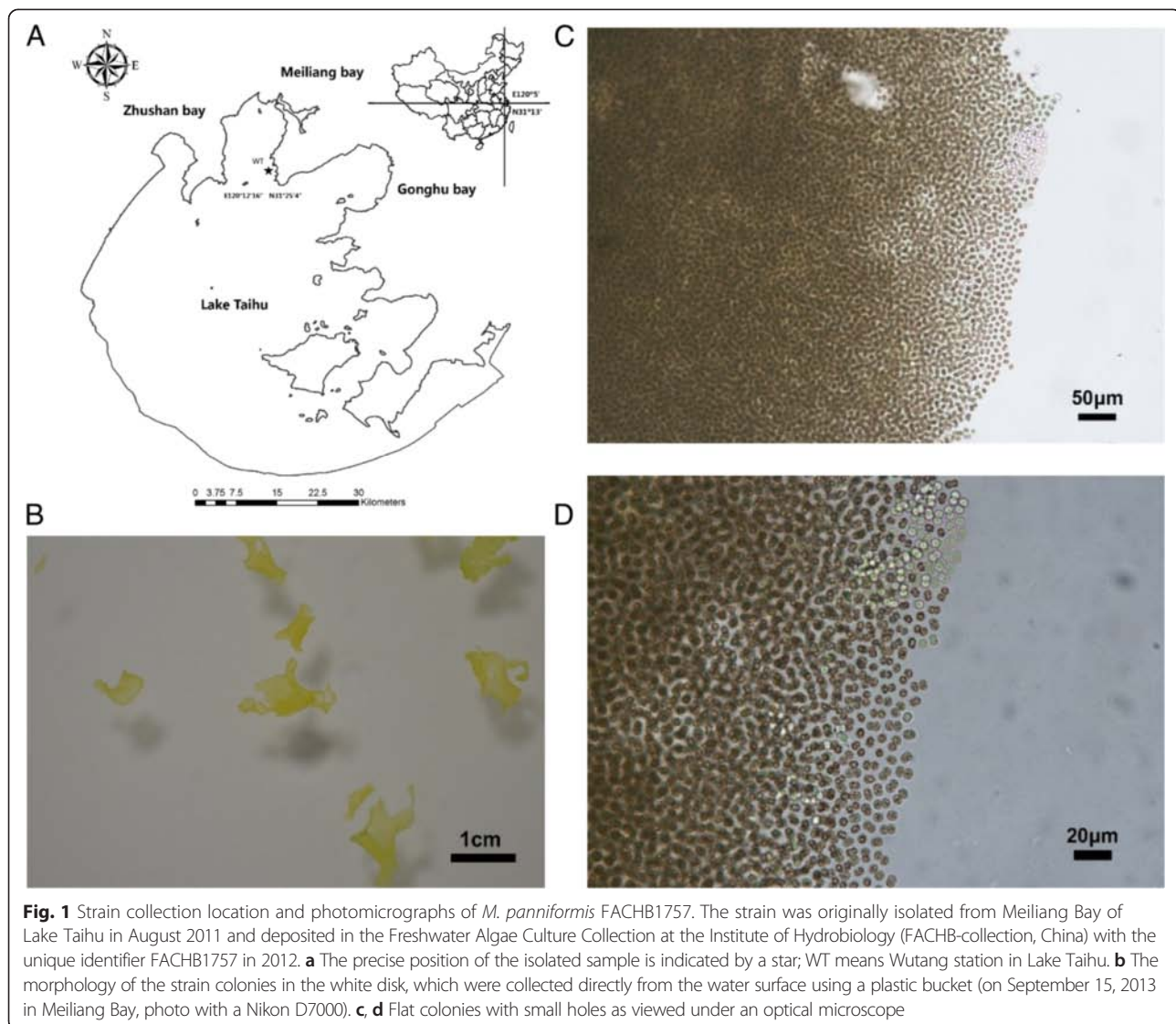
Organism information

Classification and features

A water bloom sample was collected directly from the water surface using a plastic bucket in Meiliang Bay of Lake Taihu in August 2011 (Fig. 1a). Lake Taihu (E 30° 56′–31°33′, N 119°54′–120°36′), the third largest freshwater lake in China, is located in the south of the Yangtze River Delta. The total area of the lake is 2338 km², with an average depth of 2 m and total capacity of 47.6 × 10⁸ m³. Lake Taihu is situated in the subtropical zone with a humid and semi-humid monsoon climate, and has suffered from severe eutrophication over the past

three decades. Meiliang Bay is located in the northern part of Lake Taihu (Fig. 1a), which has a surface area of 100 km², depth of 1.8–2.3 m, and is currently hypereutrophic [30].

Some *Microcystis* colonies in the sample disintegrated during the sample collection process; thus, only those macroscopic colonies with significant monolayer were collected with 3-ml pipets (BD Falcon, USA), and transferred into 50-ml centrifuge tubes (Corning, USA), and immediately shipped to the laboratory. Finally, macroscopic colonies that had flattened irregular up to monolayers with small holes (in old colonies) were identified as *M. panniformis* by examination under an optical microscope. *Microcystis panniformis* FACHB1757 was obtained, and this strain was then stored at the Freshwater Algae Culture Collection at the Institute of Hydrobiology, Chinese Academy of Sciences.



The general characteristics of *M. panniformis* FAC HB1757 are summarized in Table 1, and a phylogenetic tree based on 16S rRNA sequences is shown in Fig. 2. The spherical cells are estimated with a diameter of 2.6 to 6.8 μm (mean 4.7 μm), be densely agglomerated, and form irregular colonies with small holes. The young stages formed small clusters of cells, which were flat or circular in outline, sometimes spheroidal, and with or

Table 1 Classification and general features of *M. panniformis* FACHB1757 according to the MIGS recommendations [69]

MIGS ID	Property	Term	Evidence code ^a
	Classification	Domain <i>Bacteria</i>	TAS [70]
		Phylum <i>Cyanobacteria</i>	TAS [71, 72]
		Class	TAS [73]
		<i>Oscillatoriothycidae</i>	
		Order <i>Chroococcales</i>	TAS [73, 74]
		Family <i>Microcystaceae</i>	TAS [74]
		Genus <i>Microcystis</i>	TAS [71, 75]
		Species <i>M. panniformis</i>	TAS [15, 31]
		Strain: <i>M. panniformis</i> FACHB1757	TAS [26]
	Gram stain	Gram-negative	TAS [76]
	Cell shape	Spherical cells	TAS [15]
	Motility	Non-motile	NAS
	Sporulation	None	TAS [76]
	Temperature range	Mesophile	NAS
	Optimum temperature	29.5 °C	IDA
	pH range; Optimum	pH 7.50-9.21; pH 8.33	IDA
	Carbon source	Autotroph, heterotroph	NAS
MIGS-6	Habitat	Fresh water	NAS
MIGS-6.3	Salinity	1.0 % (maximum)	IDA
MIGS-22	Oxygen requirement	Aerobic	NAS
MIGS-15	Biotic relationship	Free-living	NAS
MIGS-14	Pathogenicity	Microcystins (MCY)	TAS [25, 77]
MIGS-4	Geographic location	Isolated Lake Taihu, China	IDA
MIGS-5	Sample collection	August, 2015	IDA
MIGS-4.1	Latitude	31.421 N	IDA
MIGS-4.2	Longitude	120.201E	IDA
MIGS-4.3	Depth	Surface 0.5 m	IDA
MIGS-4.3	Altitude	11 m	IDA

^aEvidence codes - IDA Inferred from Direct Assay, TAS Traceable Author Statement (i.e., a direct report exists in the literature), NAS Non-traceable Author Statement (i.e., not directly observed for the living, isolated sample, but based on a generally accepted property for the species, or anecdotal evidence). These evidence codes are from the Gene Ontology project [78]

without an internal hollow. The old stages formed colonies with small holes, which later disintegrated into small groups. The mucilage (margin of colonies) was diffuse, and cells did not overlap. The margin of the colonies was smooth or (in old colonies) irregular. Cell density was regular and evenly agglomerated, sometimes in indistinct rows. Diagnostic characteristics included flat colonies with small holes, toxicity, homogeneously arranged cells, and life cycle was characterized by distinct benthic and planktonic phases [15, 31]. The distribution was tropical, and this is likely a pantropical species (e.g., S. Africa, N. Australia, S. America, Africa, China, Vietnam and New Zealand) [13, 15, 16, 26, 31, 32].

Phylogenetic analysis

Whole genome comparative analysis between *M. panniformis* FACHB1757 and 13 other cyanobacterial species was performed. General information of related genome data is shown in Table S1 (Additional file 1), and all data sets were downloaded from NCBI. The main water bloom-forming cyanobacterial species in freshwater and brackish water worldwide, particularly those in the Lake Taihu region, were included. Unicellular colony-forming *Microcystis* and filamentous heterocystous *Dolichospermum* (formerly known as the planktonic *Anabaena*) were the main components of cyanobacterial blooms in Lake Taihu [33]. The *Aphanizomenon*, *Pseudanabaena*, *Cylindrospermopsis*, *Raphidiopsis*, *Planktothrix*, *Synechocystis*, and *Synechococcus* species occurred as dominant species or accompanying species in blooms of Lake Taihu (including Lake Wuli) across different seasons. Among the 14 genome sequences, 691 single-copy gene families were annotated by OrthoMCL (version 2.0.9) [34], and MEGA6 [35] was used to construct a phylogenetic tree based on these sequences (Fig. 3).

The phylogenetic tree shows that *M. panniformis* FACHB1757 and *M. aeruginosa* NIES843 shared a significantly high similarity, and there was no clear division between *M. panniformis* and *M. aeruginosa* strains in the phylogenetic tree. The *Microcystis* lineage is distinct from the lineage that contains the unicellular *Synechocystis*, *Synechococcus*, and other multicellular cyanobacteria. Furthermore, the *Synechocystis* sp. PCC 6803 genome is more closely related to *Microcystis* than other strains. This result is congruent with previously published results based on 16S rRNA sequences [36–39]. Topological relationships between species in the phylogenetic tree based on single-copy gene families were generally consistent with the phylogenetic tree based on 16S rRNA sequences (Fig. 2).

Although *Microcystis* can be identified based on 16S rRNA and single-copy gene families sequences at the genus level, taxonomy of *Microcystis* at the species level was controversial in the past few decades, and

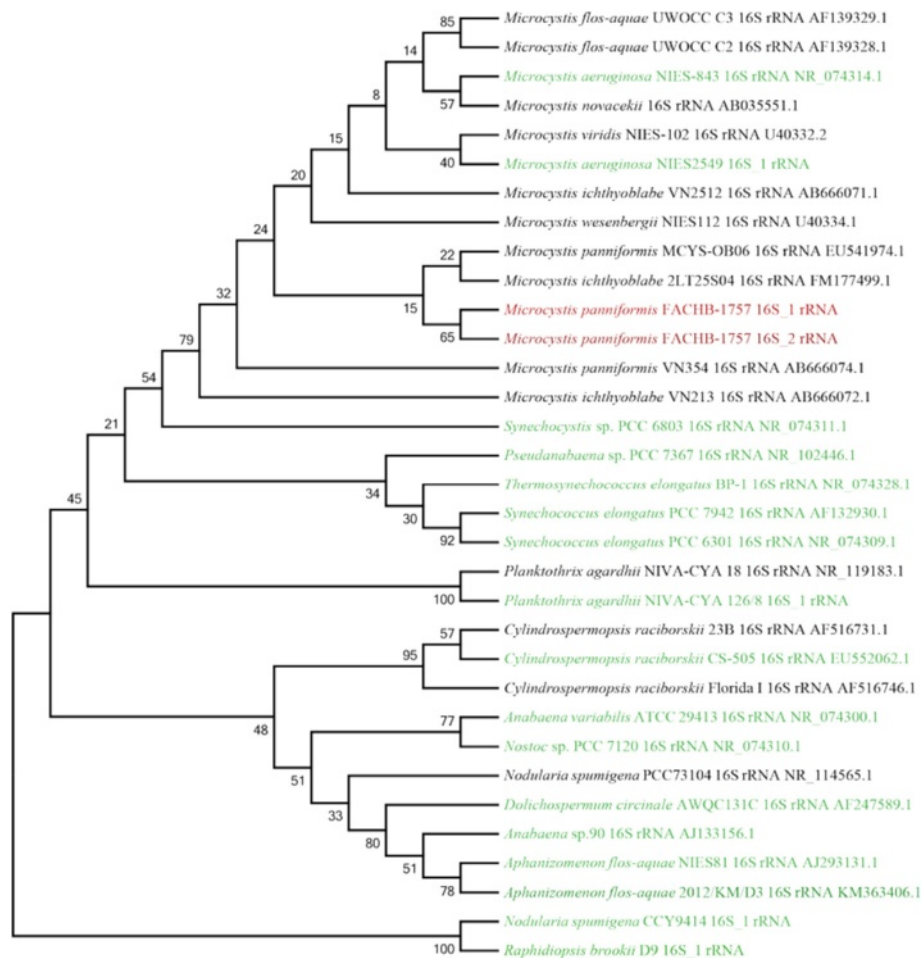


Fig. 2 Phylogenetic tree showing the position of *M. panniformis* FACHB1757. The dendrogram is based on the 16S ribosome RNA complete sequence of *M. panniformis* FACHB1757, *M. aeruginosa* NIES843, *M. aeruginosa* PCC7806, *M. aeruginosa* NIES 2549, and representatives of other cyanobacterial genera (*Synechocystis*, *Pseudanabaena*, *Synechococcus*, *Thermosynechococcus*, *Planktothrix*, *Dolichospermum*, *Anabaena*, *Cylindrospermopsis*, *Nodularia*, *Nostoc*, *Aphanizomenon*, *Raphidiopsis*) downloaded from NCBI (sequences without accession numbers were extracted from annotation files of the corresponding genomes) using the neighbor-joining algorithm with 100 bootstrap replications using MEGA6. A bootstrap consensus tree was constructed and is shown. The two copies of 16S rRNA sequences of *M. panniformis* FACHB1757 are labeled in red. The relationship between *M. panniformis* FACHB1757 and other important algae species in *Cyanophyceae* are demonstrated. Species colored in green have whole genome data available in NCBI

five species have even been unified into a single species [13]. 16S rRNA sequence estimation can be ambiguous when analyzing certain *Microcystis* species with distinct morphologies, as occurred when analyzing *M. panniformis* and *M. ichthyoblabe* (Fig. 2). Therefore, the whole reference genome sequence data was expected to play a crucial role in species classification of *Microcystis*. However, the currently available cyanobacterial genome sequences are highly limited. Only three *Microcystis* strains with complete genomic sequences are available, including *M. aeruginosa* NIES843 and NIES2549, and *M. panniformis* FACHB1757 reported here. Furthermore, the further species concepts and more useful molecular approaches should be proposed to classify the species/strain divergences in *Microcystis* [40, 41].

Genome sequencing information

Genome project history

Microcystis panniformis FACHB1757 was selected for sequencing because of its obvious morphological characteristics; in particular, the macroscopic colonies with significant monolayer can even exceed 30 mm during the summer and early autumn in Lake Taihu. More importantly, until recently, only complete genomes of *M. aeruginosa* strains (including strains NIES843 and NIES2549) have been published. The complete genome sequence of *M. panniformis* FACHB1757 would only be the third for *Microcystis*. The sample information for *M. panniformis* FACHB1757 is available in NCBI under Bio-Sample ID SAMN03392520. A DNA library with an insert size of 10 Kb was constructed, and the whole

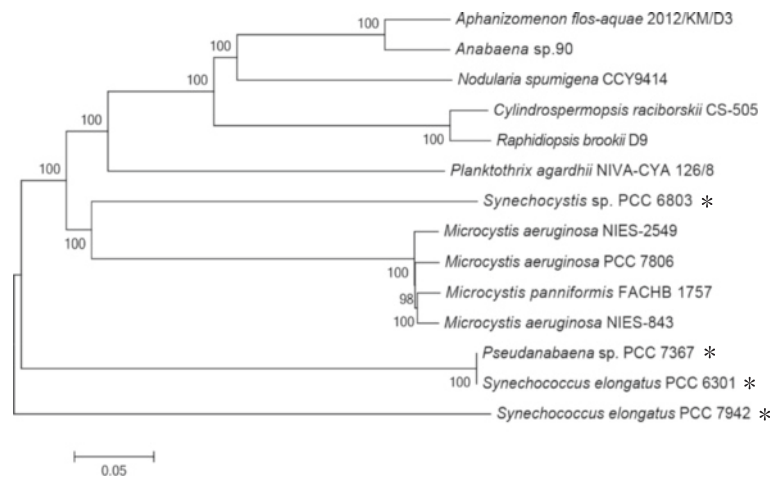


Fig. 3 Phylogenetic tree of water bloom-forming cyanobacterial species and representative cyanobacteria. The nucleotide divergence tree was constructed using the neighbor-joining algorithm based on 691 sequences of single-copy gene families annotated by OrthoMCL with 100 bootstrap replicates. The representative cyanobacteria that cannot form water blooms are indicated with an asterisk

genome was sequenced to 48-fold coverage. The completed genome sequence was assembled and uploaded to GenBank under accession number. CP011339. Project details were deposited to NCBI BioProject PRJNA277430. A summary of the project information can be found in Table 2.

Growth conditions and genomic DNA preparation

Microcystis panniformis FACHB1757 colonies collected from the field were grown in MA medium [42] and incubated in 24-well culture plates for 4 wk. Then, floating colonies were transferred to the capped tubes that contained 5 ml of MA culture medium to finally form a unialgal culture. All cultures were grown at 25 ± 1 °C with a 12 h light/12 h dark cycle under a photon

irradiance of 25 $\mu\text{mol photons}/(\text{m}^2 \cdot \text{sec})$ provided by daylight fluorescent lamps. Total genomic DNA of *M. panniformis* FACHB1757 was extracted using a commercial DNA isolation kit (DNeasy[®] Plant Mini Kit, Qiagen, USA) following the manufacturer's instructions, and analyzed by micro-volume fluorescence detection (NanoDrop[™] 8000 Spectrophotometer, Thermo Scientific, USA) and electrophoresis in 0.8 % agarose gel stained with ethidium bromide. The isolated DNA was eluted with 50 μl of the elution buffer from the commercial kit and then stored at -20 °C until subsequent analyses.

Genome sequencing and assembly

First, the genome was surveyed using an Illumina Hiseq sequencer to detect the purity of the cultured unialgal strain. The insert size of the next generation pair-end library was 100 bp, and 1 Gbp raw data was produced in total. All reads were mapped to the *M. aeruginosa* NIES843 reference complete genome, and more than 80 % of reads matched well. Subsequently, the genome was sequenced using PacBio RS II. Genomic DNA was sheared by Covaris S220 g-TUBE. A 10 Kb library was constructed using a PacBio template prep kit and sequenced using the PacBio SMRT platform. In total, two SMRT cells were run, and 303 megabase pair raw data was obtained. After filtering, the mean read length was 7143 bp with a quality of 0.84, and the longest read was 31,225 bp. HGAP (version 2.2.3) was used for genome assembly. Long reads were chosen as seeds, and the other reads were mapped to the seeds using Blasr (version 1.3.1.132871) [43] for error correction. After alignment, the accuracy of seed sequences were optimized to 99 % to meet the requirements of the Sanger assembly software. There was a total of

Table 2 Project information

MIGS ID	Property	Term
MIGS-31	Finishing quality	Complete
MIGS-28	Libraries used	2 PacBio SMRT cells
MIGS-29	Sequencing platforms	PacBio RSII
MIGS-31.2	Fold coverage	43.39
MIGS-30	Assemblers	HGAP 2.2.3
MIGS-32	Gene calling method	RAST
	Locus Tag	VL20
	GenBank ID	CP011339
	GenBank Date of Release	August 11, 2015
	GOLD ID	Gp0111943
	BIOPROJECT	PRJNA277430
MIGS-13	Source Material Identifier	FACHB1757
	Project relevance	Environmental

128 Mbp of high quality long seed reads, which had an average length of 7898 bp. Celera Assembler (version 8.1) [44] was then used to assemble the seed reads into contigs and Quiver [45] was used for second error correction. Contigs were assembled into the final complete genome sequence using minimus2 in AMOS (version 3.1.0). The final genome consisted of a complete circular 5,686,839 bp chromosome with a GC content of 42.35 % and a 38,683 bp plasmid with a 43.97 % GC content. Sequencing depths were 44.85 and 128.42, respectively.

Genome annotation

TRs were predicted by Tandem Repeat Finder (version 4.07b) [46] and Microsatellite identification tool (version 1.0), which can both identify perfect and compound micro-satellites. Prediction and annotation of the genome were done using the RAST server (version 2.0) [47]. RAST integrated tRNAscan-SE, and the *search_for_rnas* tool was used to call RNA genes across the chromosome. For gene estimation, GLIMMER2 was used to represent putative genes. Subsequently, a similar search was performed against FIGfams to identify the determined genes and annotate their functions. Moreover, all putative protein-coding genes were assigned to a category using databases including Clusters of Orthologous Groups (COG), Gene Ontology (GO), Kyoto Encyclopedia of Genes and Genomes (KEGG), Swiss-Prot, and Non-Redundant Protein Database.

Genome properties

The genome assembly contained a complete circular chromosome sequence (5.69 M) and a plasmid (38.68 K). The schematic representation of the circular chromosome of *M. panniformis* FACHB1757 was showed in Fig. 4. Related genome assembly and annotation information can be found in Table 3. Nucleotide homology search of *M. panniformis* FACHB1757 and *M. aeruginosa* NIES843 genomes was conducted by BLAST, and similarity between the two species was 83.82 % (Additional file 1: Figure S1). A total of 1944 TRs were found in the genome, including 27 microsatellites, 1742 mini-satellites, and 176 satellites. Genome statistics are shown in Table 4. In total, there were 6567 genes, which included 48 RNA genes and 6519 protein-coding genes. Among the 6519 proteins, most contained around 100 amino acids (Additional file 1: Figure S2), and by compared with function databases mentioned above, 60.15 % of them were determined to have specific functions. There were 42 tRNA genes, and two copies of the rRNA gene cluster were found in the same direction. Function assignments of 6519 putative protein-coding genes were searched against several frequently used databases mentioned above; 3260 genes were assigned to COGs, of which 235 participated in

signal transduction. Search of Pfam domains detected 3997 candidates. According to the subsystem classification results processed by RAST, 72 % of determined genes belong to specific subsystems, and the distribution of each category is demonstrated in (Additional file 1: Figure S3). The result of COG function annotation is shown in Table 4, and details of each COG cluster can be found in Additional file 2. The genes assigned to GO categories by InterProScan (version 5.4-47.0) [48] were classified into cellular components, molecular functions, and biological processes. Genes distributed in each category and their functions are shown in (Additional file 1: Figure S4). In the GO data, 309 signal function-related genes were found. KEGG matched 897 functional genes to related systems, as shown in (Additional file 1: Figure S5). Final gross function annotation outcomes are provided in (Additional file 1: Table S2).

Insights from the genome sequence

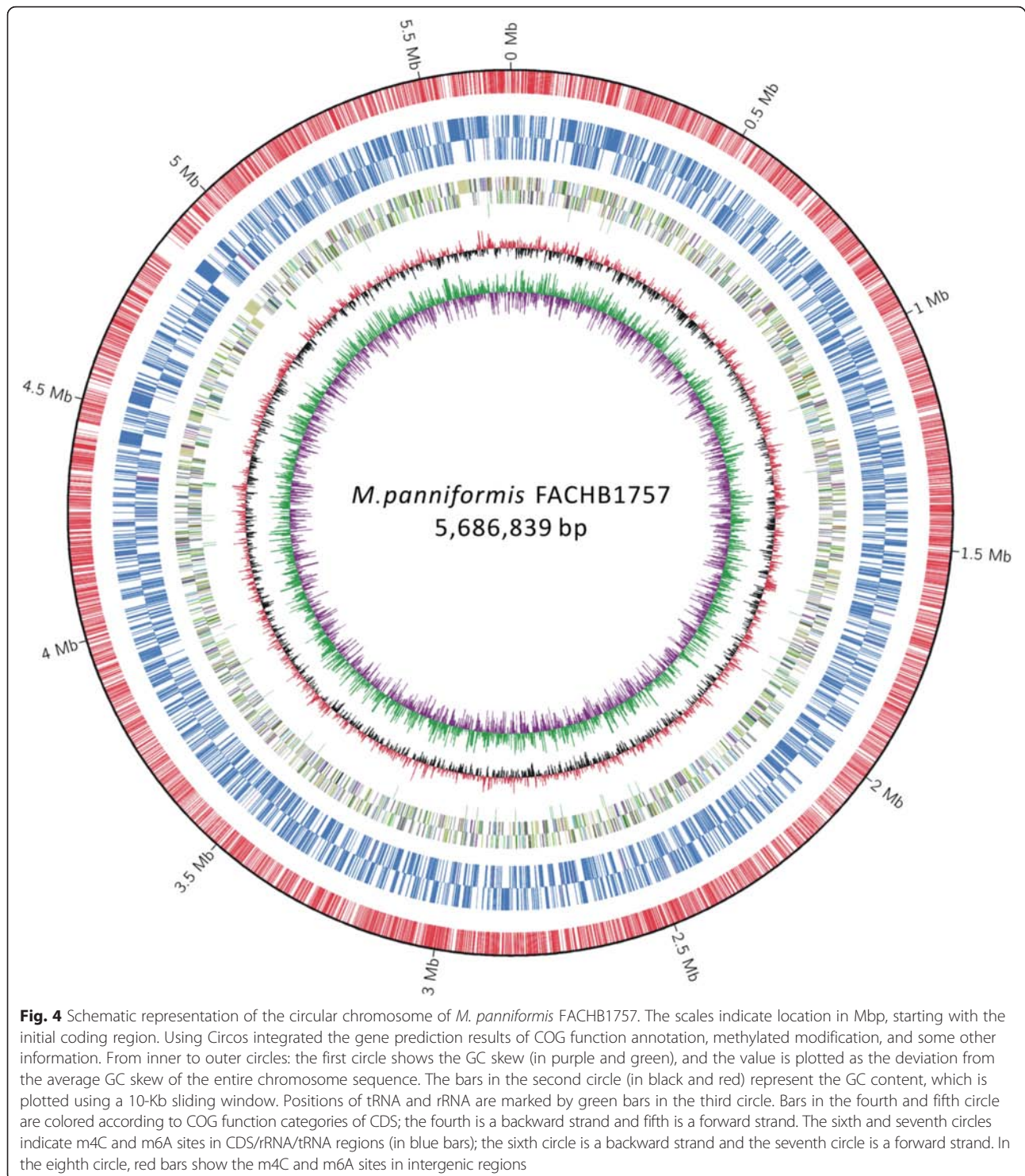
Comparative *Microcystis* species genomes

Gene ortholog analysis

Genes of four *Microcystis* species were compared (Fig. 5), and 2669 highly conserved orthologous genes were shared, which are representative of the core genome. Moreover, each genome had strain-specific genes, which varied from 296 to 1900. The *M. aeruginosa* NIES2549 genome, which has 1388 unique genes, is 1.5 Mbp smaller than that of *M. aeruginosa* NIES843, which only has 296 unique genes (*M. aeruginosa* NIES843 has 1388). *Microcystis panniformis* FACHB1757 was shown to have 1900 specific genes, which was the greatest amount among the four strains, even though its genome was not the longest.

Secondary metabolite gene clusters

Microcystin was reported to enhance colony formation in *Microcystis* spp. and plays a key role in the persistence of their colonies and the dominance of *Microcystis* [49]. As in *M. aeruginosa* NIES843 and *M. aeruginosa* PCC7806 genomes, the microcystin synthetase gene cluster (*mcyA-J*) was highly conserved in *M. panniformis* FACHB1757 from coordinates 3,496,704 to 3,541,027. Additionally, the distinct thioesterase type II coding gene *mcyT*, which occurs in toxic strains, and 4-PPT transferase (4-PPTase) were both located far from the *mcy* gene cluster at coordinates 869,702 to 869,286 and 915,377 to 916,039, respectively, which are similar to the distributions observed in *M. aeruginosa* NIES843. Notably, there was an absence of *mcnA* and *mcnB* in the *M. panniformis* FACHB1757 chromosome. *mcnA* codes polyketide biosynthesis proteins, and *mcnB* is the first open reading frame of mamestra configurata nucleopolyhedrovirus B. Together with *mcnC* and *mcnE*, these four genes compose the cyanopeptolin



synthesis gene cluster. *mcnD* was not found in the *M. panniformis* FACHB1757 genome; thus, the cyanopeptolin produced was non-halogenated and identical to that of *M. aeruginosa* NIES843 and PCC7806. Toxins may contribute to the adaptation of this strain to its specific ecological niche in eutrophic waters of tropical

and subtropical zones. In addition, a putative polyketide synthase gene cluster, which may encode additional small polypeptides found in *M. aeruginosa* NIES843 (coordinates 2,508,556–2,513,289), was detected in *M. panniformis* FACHB1757 at coordinates 4,425,371 to 4,430,104. The change in location of the genes

Table 3 Genome statistics

Attribute	Value	% of Total
Genome size (bp)	5,686,839	100.00
DNA coding (bp)	4,616,631	81.18
DNA G + C (bp)	2,408,639	42.35
DNA scaffolds	1	100.00
Total genes	6,567	100.00
Protein coding genes	6,519	99.27
RNA genes	48	0.73
rRNA genes	6	0.09
tRNA genes	42	0.64
Pseudo genes	-	-
Genes in internal clusters	-	-
Genes with function prediction	3,921	100.00
Genes assigned to COGs	3,373	86.02
Genes with Pfam domains	2,067	52.72
Genes with signal peptides	309	7.88
CRISPR repeats	3	-
Genes with transmembrane helices	-	-

mentioned above reflected the extensive structural variation between *M. panniformis* FACHB1757 and *M. aeruginosa* NIES843.

Conserved gene clusters

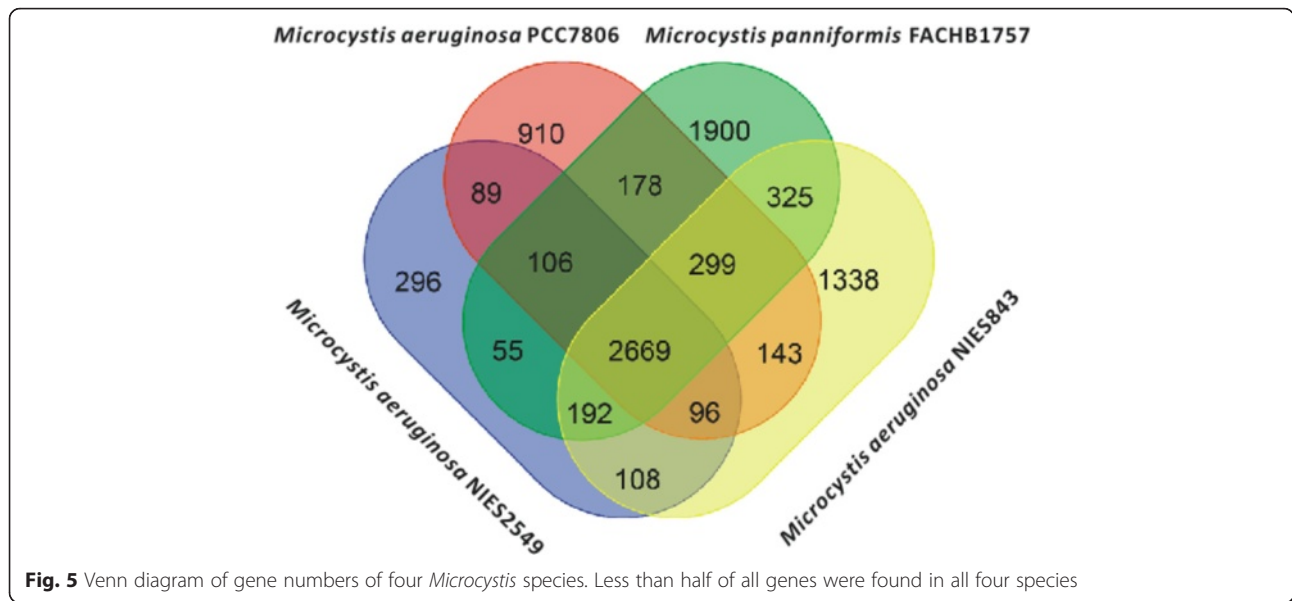
Four functional clusters of conserved genes related to microcystin synthesis, colony formation, photo-regulation, and nutrient assimilation were also compared among the four *Microcystis* strains. In the microcystin synthesis gene cluster, the *mcy* and *mcn* gene clusters were not found in *M. aeruginosa* NIES2549. This is consistent with the results of a previous study, which showed that *M. aeruginosa* NIES2549 is a nontoxic strain [50]. With regard to colony formation, *M. aeruginosa*, *M. wesenbergii*, and *M. panniformis* all have typical macroscopic colony structure when observed by naked eye in Lake Taihu during summer and autumn water blooms. *Microcystis panniformis* seems to be the largest, and can even have more than 30 mm colonies. Polysaccharides and microcystin play important roles in the process of *Microcystis* colony formation. The maximum EPS content was found in *M. wesenbergii* and *M. aeruginosa*, which are not the largest and are only approximately 100 μm [51], but positive correlations between EPS and *Microcystis* colony size in cultures were supported by previous studies [52–54]. *mrpC* and *epsL* were absent from all four strains, and only *M. aeruginosa* NIES843 contained *cpsF*, although *tagH*, *capD*, *csaB*, and *rfbB* were conserved

Table 4 Number of genes associated with general COG functional categories

Code	Value	% Age	Description
J	165	2.51	Translation, ribosomal structure and biogenesis
A	0	0.00	RNA processing and modification
K	131	1.99	Transcription
L	620	9.44	Replication, recombination and repair
B	1	0.02	Chromatin structure and dynamics
D	47	0.72	Cell cycle control, Cell division, chromosome partitioning
V	67	1.02	Defense mechanisms
T	138	2.10	Signal transduction mechanisms
M	203	3.09	Cell wall/membrane biogenesis
N	30	0.46	Cell motility
U	52	0.79	Intracellular trafficking and secretion
O	152	2.31	Posttranslational modification, protein turnover, chaperones
C	185	2.82	Energy production and conversion
G	131	1.99	Carbohydrate transport and metabolism
E	215	3.27	Amino acid transport and metabolism
F	62	0.94	Nucleotide transport and metabolism
H	146	2.22	Coenzyme transport and metabolism
I	65	0.99	Lipid transport and metabolism
P	188	2.86	Inorganic ion transport and metabolism
Q	126	1.92	Secondary metabolites biosynthesis, transport and catabolism
R	535	8.14	General function prediction only
S	483	7.35	Function unknown
-	2,826	43.02	Not in COGs

The total is based on the total number of protein-coding genes in the genome

in all four strains. Furthermore, *mvn* codes for a lectin in *M. panniformis* FACHB1757 and *M. aeruginosa* PCC7806, which specifically binds to a sugar moiety present on the surface of *Microcystis* cells. Additionally, a binding partner of MVN was identified in the lipopolysaccharide fraction of *M. aeruginosa* PCC7806, which involved in the *Microcystis* colony formation [55]. Together, the toxin-, EPS-, and lectin-related genes may explain the reason why *M. panniformis* FACHB1757 usually aggregates and produces a larger colony in Lake Taihu during water blooms. In the photo-regulation cluster, *psb*, *apc* and *gvp* with the exception of *gvpC* were all detected. It is interesting that *gvpC* is absent from *M. panniformis* FACHB1757, because this gene encodes GvpC, which is a highly conserved expressed protein in some genera that is closely related to gas vesicles [56–58]. Genes related to nutrient assimilation include *ntc*, *pst*, and *sph* clusters. *ntcB*, *pstA*, *pstB1*, *pstB2*, and *pstC* were only absent from *M. aeruginosa* PCC7806 among the four strains, which



may be accounted for by the incompleteness of the strains genome. Detailed information about function and coordinates of each gene are shown in (Additional file 1: Table S3).

Genome structure and constitution comparison

The genomes of *M. aeruginosa* NIES843 and *M. aeruginosa* NIES2549 have no plasmids, whereas a 38 Kb plasmid with a 43.97 % GC content was detected in *M. panniformis* FACHB1757 in this study. The stable presence of plasmids may play an important role in some *Microcystis* obtaining competitive advantages [59–61]. *Microcystis aeruginosa* NIES-843 is the first strain of the genus *Microcystis* to be sequenced for its complete genome with the ABI 3770xl sequencer. Since then, the second completed *Microcystis* genome (of *M. aeruginosa* NIES-2549) was released on the April 29, 2015. Thus, the whole genome at the nucleic acid

level was compared between *M. panniformis* FACHB 1757 and *M. aeruginosa* NIES-843. Mauve, which was designed for identification and alignment of conserved genomic sequences with rearrangements and horizontal transfer, was used to conduct comparative genomic sequence analysis [62]. As shown in (Additional file 1: Figure S1), *M. panniformis* FACHB1757 underwent extensive chromosome structure rearrangement, which indicates that *Microcystis* genomes are highly plastic [36].

Self-defense system

Restriction modification system

Comparison with REBASE [63], a restriction enzyme database containing information about restriction enzymes, revealed that DNA methyltransferases and related proteins are involved in the biological process of R–M, and

Table 5 Sequence structure and general information of motifs in the whole genome

Motif	Modified position	Type	Motifs detected	# of motifs detected	# of motifs in genome	Mean QV	Mean motif coverage	Partner motif
GATC	2	m6A	78.55 %	37,874	48,218	45.44	22.91	GATC
GAATTC	3	m6A	74.54 %	1938	2,600	44.25	22.53	GAATTC
GCTGDAG	6	m6A	73.70 %	995	1,350	43.72	22.95	-
GGTGGA	6	m6A	70.96 %	1935	2,727	43.22	22.81	-
GACGNAC	6	m6A	70.26 %	723	1,029	42.58	23.11	-
ACCACC	4	m6A	69.67 %	2410	3,459	41.91	22.82	-
CAAGNNNNNTTTC	3	m6A	69.02 %	176	255	41.45	21.48	-
GATATC	2	m6A	67.42 %	2055	3,048	42.29	23.09	GATATC
MCGRAG	5	m6A	52.23 %	3390	6,491	41.64	22.35	-
GCWGC	2	m4C	24.17 %	3911	16,184	37.52	25.13	GCWGC
RGATCY	5	m4C	19.09 %	808	4,232	36.99	25.80	RGATCY
GGCC	3	m4C	18.02 %	3721	20,654	37.67	26.39	GGCC

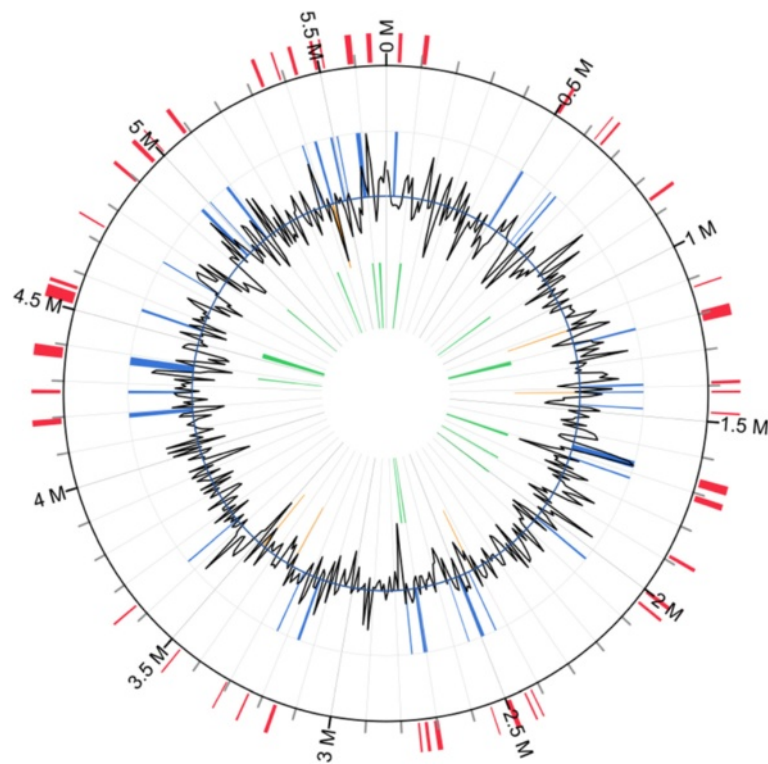


Fig. 6 GEIs distribution in the chromosome of *M. panniformis* FACHB1757. From inside to outside, green bars illustrate IslandPick prediction, orange bars show the results annotated by SIGI-HMM, and blue bars are predicted by IslandPath-DIMOB. Red bars indicate integrated GEIs candidate positions. Black line plot around the small circle reveal the GC content

277 restriction enzymes were found. Detailed classification revealed that 12 and 130 enzymes belonged to type I and type II systems, respectively, which together represented 46.93 % of all enzymes, and are categories of rapidly evolving genes [64]. Sixty-three, 10, and 2 enzymes, respectively, belonged to type IIG, type III, and type IV systems, and one control protein restriction enzyme and 58 unknown enzymes were also found.

Methylation modification analysis

It is widely thought that methylation modification is associated with R-M systems and participates in self-defense against foreign genome invasion. Genome methylation modification and methyl-transferase recognition sequence motifs were analyzed using SMRT (version 2.3.0). In the chromosome, 3204 m4C (N^4 -methylcytosine), 9,758 m6A (N^6 -methyladenine), and 31,845 other

Table 6 Functions and types of all 36 GEIs in chromosome

Function	Advantage conferred	GEI type	Related GEIs
Alkaline phosphatase	Increased metabolic versatility	Metabolic	GEI2,GEI7,GEI10,GEI17,GEI19,GEI23,GEI26,GEI27,
Toxin/Antitoxin protein	Competitiveness	Pathogenicity, resistance	GEI1,GEI6,GEI13,GEI31,GEI32,GEI33
Transferase	Increased metabolic versatility	Metabolic	GEI4,GEI9,GEI15,GEI21,GEI24,GEI25,GEI30,GEI36
Transposase	Increased metabolic versatility	Metabolic	GEI1,GEI3,GEI4,GEI11,GEI12,GEI15,GEI16,GEI18, GEI24,GEI29,GEI34
Hat/HatR	Increased metabolic versatility, increased adaptability	Fitness	GEI11,GEI28
Heat shock protein	Increased metabolic versatility, increased adaptability	Synthesis, fitness	GEI31
PsaE	Increased metabolic versatility	Metabolic, fitness	GEI9

modified bases were marked as modified (details are available in Additional file 3). Corresponding motif information is included in Table 5.

CRISPR system

MinCED derived from the CRT [65], was used to predict CRISPR structure. CRISPR are extensively found in prokaryotes and are thought to compose a CRISPR-associated system, which is a putative immune system based on RNA-interference [66]. Three candidate CRISPR clusters on chromosome sequence were annotated under strict parameter and 1 CRISPR on plasmid (further information is available in Additional file 4).

Genomic islands

GEIs are particularly influential in microorganism genomes with regard to virulence, antibiotic resistance, metabolic, symbiosis, or other important adaptations [67]. GEIs have substantial roles in horizontal gene transfer, which is now widely acknowledged as an important force that shapes bacterial genome structure. Island Viewer (version 2.0) [68] was used to predict the GEIs in *M. panniformis* FACHB1757. Island Viewer integrates SIGI-HMM, Island Pick, and Island Path-DIMOB and built-in databases, including the Virulence Factor Database and Antibiotic Resistance Gene Database. Thirty-six GEIs were found using Island Viewer, and their positions are shown in Fig. 6. Different kinds of functions were identified and are summarized in Table 6. Transposases were identified in most of the GEIs, as they participated in horizontal gene transfer. Toxin-related gene clusters were annotated in six GEIs and probably affect competitiveness and fitness. Some functional genes, such as *hat/hatR*, were also detected, which indicates the enhanced adaptability and metabolic versatility in this strain.

Conclusions

This study presents the complete whole genome sequence of a newly recorded species in China, *M. panniformis*, and demonstrates several genomic perspectives, including comparison with nine other water bloom-forming cyanobacterial species. A 5.6 Mbp chromosome with a 38 Kbp plasmid was reported, and gene function, methylation modification, CRISPR, and GEIs throughout the genome were described. Large-scale of structure variation was demonstrated by comparison with *M. aeruginosa* genomes. A Venn diagram of four *Microcystis* strains showed gene quantity and category variation as a result of evolutionary divergence and revealed that *Microcystis* has undergone flexible genome evolution.

Additional files

Additional file 1: Tables S1-3 and Figures S1-5. Table S1. General information regarding phylogenomics and comparative genomic analysis.

Table S2. Function annotation assignment from different databases.

Table S3. Conserved gene cluster function and position in the genomes of four *Microcystis* strains. Figure S1. Whole genome comparison with *M. aeruginosa* NIES843 at the nucleic acid level. Large-scale structural rearrangement was obvious at the nucleic acid level. Blocks on the line for *M. panniformis* FACHB1757 show that the sequences of the two species were in the same direction, whereas blocks under the line indicated that the sequences ran in the opposite direction. Figure S2. Length distribution of annotated protein-coding genes. The majority of fractions varied in length from 100 to 150. Figure S3. Subsystems and category distributions of genes. Functional annotation by RAST. Figure S4. Histogram demonstrating functional annotation results against GO. Figure S5. Number of each category in function annotation results against KEGG. (DOCX 599 kb)

Additional file 2: The COG function annotation in each COG cluster. (XLSX 114 kb)

Additional file 3: N⁴-methylcytosine, N⁶-methyladenine, and other modified bases throughout the whole chromosome. (XLSX 7893 kb)

Additional file 4: Annotation of three candidate CRISPR clusters on the chromosome sequence and one CRISPR cluster on the plasmid using strict parameters. (XLSX 14 kb)

Abbreviations

CRISPR: Clustered Regularly Interspaced Short Palindromic Repeats; CRT: CRISPR Recognition Tool; EPS: Extracellular polysaccharide; GEIs: Genome islands; MinCED: Mining CRISPRs in Environmental Datasets; R-M: Restriction-modification; TRs: Tandem repeats.

Competing interests

The authors declare that they have no competing interests.

Authors' contributions

J-YZ and RG contributed equally to this paper. J-YZ carried out microscopy, sample collection and purification, and drafted the manuscript. RG carried out genomic perspective analysis, the genomic evolution analysis, and conceived the manuscript. HL participated in sample culture and genomic DNA extraction. PX provided information about the *M. aeruginosa* PCC7806 genome. LD and D-MC participated in genome assembly and annotation. H-JZ and B-CZ contributed technical assistance to sample collection and purification. HL, PX and G-LY helped with discussions. Professor Z-HL and R-HL conceived the study, and its design and coordination, and helped draft the manuscript. All authors read and approved the final manuscript.

Acknowledgements

This work was performed under the authority of Wuxi Environmental Monitoring Center, the State Key Laboratory of Bioelectronics (Southeast University), and Institute of Hydrobiology (The Chinese Academy of Sciences). It was partially supported by Major Science and Technology Program for Water Pollution Control and Treatment (2012ZX07101-013-06) to HZ, the National Natural Science Foundation of China (No.61227803), and the NSFC-JST Project (No.41261140337). We thank Ting Song for mapping the collected strain location (Fig. 1a). We gratefully acknowledge the help of Mr. Ge (Director of Wuxi Municipal Environmental Protection Bureau, China) and Mr. Li (Deputy Director of Wuxi Municipal Environmental Protection Bureau, China) in the aspect of administrative support for this project.

Author details

¹State Key Laboratory of Bioelectronics, School of Biological Sciences and Medical Engineering, Southeast University, Nanjing 210096, China. ²Wuxi Environmental Monitoring Center, Wuxi 214121, China. ³Institute of Hydrobiology, The Chinese Academy of Sciences, Wuhan, Hubei 430072, China. ⁴Shenzhen Key Laboratory for Marine Bio-resource and Eco-environment, College of Life Sciences, Shenzhen University, Shenzhen 518060, China. ⁵Nextomics Biosciences Co., Ltd., Wuhan, Hubei 430072,

China. ⁶Department of Biomedical Engineering, College of Engineering, Peking University, Beijing 100871, China.

Received: 4 September 2015 Accepted: 5 January 2016

References

- Bister B, Keller S, Baumann HI, Nicholson G, Weist S, Jung G, et al. Cyanopeptolin 963A, a Chymotrypsin Inhibitor of *Microcystis* PCC 7806. *J Nat Prod*. 2004;67:1755–7. PubMed <http://dx.doi.org/10.1021/np049828f>.
- Kardinaal WEA, Visser PM. Dynamics of cyanobacterial toxins. In: Codd G, Lindsay J, Young F, Morrison L, Metcalfe J, editors. *Harmful Cyanobacteria*. Dordrecht, The Netherlands; 2005.p.25–40. http://dx.doi.org/10.1007/1-4020-3022-3_1.
- Portmann C, Blom JF, Gademann K, Jüttner F. Aerucyclamides A and B: Isolation and Synthesis of Toxic Ribosomal Heterocyclic Peptides from the Cyanobacterium *Microcystis aeruginosa* PCC 7806. *Journal of Natural Products*. 2008;71:1193–1196. PubMed <http://dx.doi.org/10.1021/np800118g>.
- Whitton BA. *Microcystis*. In: Šejnohová L, Maršálek, editors. *Ecology of Cyanobacteria II: Their Diversity in Space and Time*. Dordrecht, The Netherlands; 2012.p.195–228. <http://dx.doi.org/10.1007/978-94-007-3855-3>.
- Paerl HW, Huismann J. Climate Blooms like it hot. *Science*. 2008;320:57–8. PubMed <http://dx.doi.org/10.1126/science.1155398>.
- Weirich CA, Miller TR. Freshwater harmful algal blooms: toxins and children's health. *Curr Probl Pediatr Adolesc Health Care*. 2014;44:2–24. PubMed <http://dx.doi.org/10.1016/j.cppeds.2013.10.007>.
- Trevino-Garrison I, DeMent J, Ahmed FS, Haines-Lieber P, Langer T, Ménager H, et al. Human illnesses and animal deaths associated with freshwater harmful algal blooms-kansas. *Toxins*. 2015;7:353. PubMed <http://dx.doi.org/10.3390/toxins7020353>.
- Jason S, Hayes PK. Molecular axonomy of Harmful Algae. In: Granéli E, Turner JT, editors. *Ecology of Harmful Algae*. Dordrecht, The Netherlands; 2006. p. 15–16. <http://dx.doi.org/10.1007/978-3-540-32210-8>.
- Sanchis D, Padilla C, Del Campo FF, Quesada A, Sanz-Alferez S. Phylogenetic and morphological analyses of *Microcystis* strains (Cyanophyta/Cyanobacteria) from a Spanish water reservoir. *NOVA HEDWIGIA*. 2005;81:431–48. <http://dx.doi.org/10.1127/0029-5035/2005/0081-0431>.
- Kondo R, Yoshida T, Yuki Y, Hiroishi S. DNA-DNA reassociation among a bloom-forming cyanobacterial genus, *Microcystis*. *Int J Syst Evol Microbiol*. 2000;50:767–70. PubMed <http://dx.doi.org/10.1099/00207713-50-2-767>.
- Otsuka S, Suda S, Shibata S, Oyaizu H, Matsumoto S, Watanabe MM. A proposal for the unification of five species of the cyanobacterial genus *Microcystis* Kützing ex Lemmermann 1907 under the rules of the Bacteriological Code. *Int J Syst Evol Microbiol*. 2001;51:873–9. PubMed <http://dx.doi.org/10.1099/00207713-51-3-873>.
- Otsuka S, Suda S, Li R, Watanabe M, Oyaizu H, Matsumoto S, et al. Phylogenetic relationships between toxic and non-toxic strains of the genus *Microcystis* based on 16S to 23S internal transcribed spacer sequence. *FEMS Microbiol Lett*. 1999;172:15–21. PubMed <http://dx.doi.org/10.1111/j.1574-6968.1999.tb13443.x>.
- Le Ai NV, Tanabe Y, Matsuura H, Kaya K, Watanabe MM. Morphological, biochemical and phylogenetic assessments of water-bloom-forming tropical morphospecies of *Microcystis* (Chroococcales, Cyanobacteria). *Phycol Res*. 2012;60:208–22. <http://dx.doi.org/10.1111/j.1440-1835.2012.00650.x>.
- Bittencourt-Oliveira MC, Moura AN, Gouvêa-Barros S, Pinto E. HIP1 DNA fingerprinting in *Microcystis panniformis* (Chroococcales, Cyanobacteria). *Phycologia*. 2007;46:3–9. <http://dx.doi.org/10.2216/06-01.1>.
- Komárek J, Komárková-Legnerová J, Sant'Anna CL, Azevedo MTDP, Senna PAC. Two common *Microcystis* species (Chroococcales, Cyanobacteria) from tropical America, including *M. panniformis* sp. nov. *Cryptogamie. Algologie*. 2002;23:159–77.
- White SH, Fabbro LD, Duivenvoorden LJ. Changes in cyanoprokaryote populations, *Microcystis* morphology, and microcystin concentrations in Lake Elphinstone (Central Queensland, Australia). *Environ Toxicol*. 2003;18:403–12. PubMed <http://dx.doi.org/10.1002/tox.10142>.
- Sant'Anna CL, Maria Teresa de PA, Pedro Américo CS, Komárek J, Komárková J. Planktic Cyanobacteria from São Paulo State, Brazil: Chroococcales Cianobactérias planctônicas do Estado de São Paulo, Brasil: Chroococcales. *Revista Brasileira de Botânica*. 2004; 27:213–227. <http://dx.doi.org/10.1590/S0100-84042004000200002>.
- Nguyen LTT, Cronberg G, Larsen J, Moestrup O. Planktic cyanobacteria from freshwater localities in Thuathien-Hue Province, Vietnam I. Morphology and distribution. *NOVA HEDWIGIA*. 2007;85:1–34. <http://dx.doi.org/10.1127/0029-5035/2007/0085-0001>.
- Borges JT, Sparrapan R, Guimarães JR, Eberlin MN, Augusti R. Chloroform formation by chlorination of aqueous algae suspensions: online monitoring via membrane introduction mass spectrometry. *J Brazilian Chem Soc*. 2008; 19:950–5. <http://dx.doi.org/10.1590/S0103-50532008000500021>.
- Luciana Retz de C, Pipole F, Werner VR, Haywood Dail Laughinghouse IV, Antonio Carlos MC, Rangel M, et al. A toxic cyanobacterial bloom in an urban coastal lake, Rio Grande do Sul state, Southern Brazil. *Brazilian J Microbiol*. 2008;39:761–9. PubMed <http://dx.doi.org/10.1590/S1517-838220080004000031>.
- Bittencourt-Oliveira MC, Cunha MCC, Moura AN. Genetic polymorphism in Brazilian *Microcystis* spp. (Cyanobacteria) toxic and non-toxic through RFLP-PCR of the *cpcBA*-IGS. *Brazilian Archives Biol Technol*. 2009;52:901–9. <http://dx.doi.org/10.1590/S1516-89132009000400014>.
- Caliskan M. Genetic Diversity of Brazilian Cyanobacteria Revealed by Phylogenetic Analysis. In: Bittencourt-Oliveira M, Piccin-Santos V, editors. *Genetic Diversity in Microorganisms*. 2012.p. 275–290. <http://dx.doi.org/10.5772/33137>.
- Edvani MS, Cícero TSG, Catarina PSR, Sílvia MLB. Ocorrência de Cianobactérias no Reservatório de Mundaú, no Estado de Pernambuco, no Período de Janeiro de 2010 a Novembro de 2011 / Occurrence of Cyanobacteria in Mundaú Reservoir, in Pernambuco State During January 2010 to November 2011. *Vigilância Sanitária em Debate: Sociedade*. 2013;1:35–42. <http://dx.doi.org/10.3395/vd.v1i3.40>.
- Bittencourt-Oliveira MC, Chia MA, de Oliveira HS B, Araujo MKC, Molica RJR, Dias CTS. Allelopathic interactions between microcystin-producing and non-microcystin-producing cyanobacteria and green microalgae: implications for microcystins production. *J Appl Phycol*. 2015;27:275–84. <http://dx.doi.org/10.1007/s10811-014-0326-2>.
- Bittencourt-Oliveira MC, Kujbida P, Cardozo KHM, Carvalho VM, Moura AN, Colepicolo P, et al. A novel rhythm of microcystin biosynthesis is described in the cyanobacterium *Microcystis panniformis* Komárek et al. *Biochem Biophys Res Commun*. 2005;326:687–94. PubMed <http://dx.doi.org/10.1016/j.bbrc.2004.11.091>.
- Zhang JY, Zhu BC, Wu ZJ, Xu T, Lu ZH. *Microcystis panniformis*-A newly recorded species of *Microcystis* in China. *J Lake Sci*. 2012;24:647–50.
- Paerl HW, Paul VJ. Climate change: links to global expansion of harmful cyanobacteria. *Water Res*. 2012;46:1349–63. PubMed <http://dx.doi.org/10.1016/j.watres.2011.08.002>.
- Paerl HW. Mitigating harmful cyanobacterial blooms in a human-and climatically-impacted world. *Life (Basel, Switzerland)*. 2014; 4:988–1012. PubMed <http://dx.doi.org/10.3390/life4040988>.
- Qin B, Zhu G, Gao G, Zhang Y, Li W, Paerl HW, et al. A drinking water crisis in Lake Taihu, China: linkage to climatic variability and lake management. *Environ Manage*. 2010;45:105–12. PubMed <http://dx.doi.org/10.1007/s00267-009-9393-6>.
- Qin B, Xu P, Wu Q, Luo L, Zhang Y. Environmental issues of Lake Taihu China. *Hydrobiologia*. 2007;581:3–14. <http://dx.doi.org/10.1007/s10750-006-0521-5>.
- Komárek J, Komárková J. Review of the European *Microcystis*-morphospecies (Cyanoprokaryotes) from nature. *Czech Phycology, Olomouc*. 2002;2:1–24.
- Wood SA, Crowe ALM, Ruck JG, Wear RG. New records of planktonic cyanobacteria in New Zealand freshwaters. *New Zealand J Botany*. 2005;43:479–92. <http://dx.doi.org/10.1080/0028825X.2005.9512969>.
- Liu Y, Xu Y, Xiao P, Pan Q, Yu G, Li R. Genetic analysis on *Dolichospermum* (Cyanobacteria; sensu *Anabaena*) populations based on the culture-independent clone libraries revealed the dominant genotypes existing in Lake Taihu, China. *Harmful Algae*. 2014;31:76–81. <http://dx.doi.org/10.1016/j.hal.2013.09.012>.
- Li L, Stoeckert Jr CJ, Roos DS. OrthoMCL: identification of ortholog groups for eukaryotic genomes. *Genome Res*. 2003;13:2178–89. PubMed <http://dx.doi.org/10.1101/gr.1224503>.
- Tamura K, Stecher G, Peterson D, Filipiński A, Kumar S. MEGA6: Molecular Evolutionary Genetics Analysis version 6.0. *Mol Biol Evol*. 2013;30:2725–9. PubMed <http://dx.doi.org/10.1093/molbev/mst197>.
- Frangeul L, Quillardet P, Castets AM, Humbert JF, Matthijs HC, Cortez D, et al. Highly plastic genome of *Microcystis aeruginosa* PCC 7806, a ubiquitous toxic freshwater cyanobacterium. *BMC Genom*. 2008;9:274. PubMed <http://dx.doi.org/10.1186/1471-2164-9-274>.
- Rantala A, Fewer DP, Hisbergues M, Rouhiainen L, Vaitomaa J, Borner T, et al. Phylogenetic evidence for the early evolution of microcystin synthesis. *Proc Natl Acad Sci U S A*. 2004;101:568–73. PubMed <http://dx.doi.org/10.1073/pnas.0304489101>.

38. Schirrmeyer BE, De Vos JM, Antonelli A, Bagheri HC. Evolution of multicellularity coincided with increased diversification of cyanobacteria and the Great Oxidation Event. *Proc Natl Acad Sci U S A*. 2013;110:1791–6. PubMed <http://dx.doi.org/10.1073/pnas.1209927110>.
39. Schirrmeyer BE, Antonelli A, Bagheri HC. The origin of multicellularity in cyanobacteria. *BMC Evol Biol*. 2011;11:45. PubMed <http://dx.doi.org/10.1186/1471-2148-11-45>.
40. Palinska KA, Surosz W. Taxonomy of cyanobacteria: a contribution to consensus approach. *Hydrobiologia*. 2014;740:1–11. <http://dx.doi.org/10.1007/s10750-014-1971-9>.
41. Dvořák P, Pouličková A, Hašler P, Belli M, Casamatta D, Papini A. Species concepts and speciation factors in cyanobacteria, with connection to the problems of diversity and classification. *Biodivers Conserv*. 2015;24:1–19. <http://dx.doi.org/10.1007/s10531-015-0888-6>.
42. Kasai F, Kawachi M, Erata M, Watanabe MM. NIES-Collection, List of Strains: Microalgae and Protozoa. In: Kasai, F., Kawachi, M., Erata, M., Watanabe, M.M., editors. National Institute for Environmental Studies. Japan, Tsukuba; 2004.P. 257.
43. Chaisson MJ, Tesler G. Mapping single molecule sequencing reads using basic local alignment with successive refinement (BLASR): application and theory. *BMC Bioinformatics*. 2012;13:238. PubMed <http://dx.doi.org/10.1186/1471-2105-13-238>.
44. Miller JR, Delcher AL, Koren S, Venter E, Walenz BP, Brownley AC, et al. Aggressive assembly of pyrosequencing reads with mates. *BMC Bioinformatics*. 2008;24:2818–24. PubMed <http://dx.doi.org/10.1093/bioinformatics/btn548>.
45. Chin CS, Alexander DH, Marks P, Klammer AA, Drake J, Heiner C, et al. Nonhybrid, finished microbial genome assemblies from long-read SMRT sequencing data. *Nat Methods*. 2013;10:563–9. PubMed <http://dx.doi.org/10.1038/nmeth.2474>.
46. Benson G. Tandem repeats finder: a program to analyze DNA sequences. *Nucleic Acids Res*. 1999;27:573–80. PubMed <http://dx.doi.org/10.1093/nar/27.2.573>.
47. Aziz RK, Bartels D, Best AA, DeJongh M, Disz T, Edwards RA, et al. The RAST Server: rapid annotations using subsystems technology. *BMC Genomics*. 2008;9:75. PubMed <http://dx.doi.org/10.1186/1471-2164-9-75>.
48. Jones P, Binns D, Chang HY, Fraser M, Li W, McAnulla C, et al. InterProScan 5: genome-scale protein function classification. *Bioinformatics (Oxford, England)*. 2014; 30:1236–1240. PubMed <http://dx.doi.org/10.1093/bioinformatics/btu031>.
49. Gan N, Xiao Y, Zhu L, Wu Z, Liu J, Hu C, et al. The role of microcystins in maintaining colonies of bloom-forming *Microcystis* spp. *Environ Microbiol*. 2012;14:730–42. PubMed <http://dx.doi.org/10.1111/j.1462-2920.2011.02624.x>.
50. Tanabe Y, Watanabe MM. Local expansion of a panmictic lineage of water bloom-forming cyanobacterium *Microcystis aeruginosa*. *PLoS One*. 2011;6:e17085. PubMed <http://dx.doi.org/10.1371/journal.pone.0017085>.
51. Zhu W, Dai X, Li M. Relationship between extracellular polysaccharide (EPS) content and colony size of *Microcystis* is colonial morphology dependent. *Biochem Sys Ecol*. 2014;55:346–50. <http://dx.doi.org/10.1016/j.bse.2014.04.009>.
52. Zhao L, Lu L, Li M, Xu Z, Zhu W. Effects of Ca and Mg levels on colony formation and EPS content of cultured *M. aeruginosa*. *Procedia Environmental Sciences*. 2011; 10, Part B:1452–1458. <http://dx.doi.org/10.1016/j.proenv.2011.09.232>.
53. Liu Y, Wang W, Geng L, Chen Y, Yang Z. Polysaccharide content and morphology of *Microcystis aeruginosa* in response to changes in metabolic carbon flux. *Fresenius Environ Bull*. 2011;20:1046–50.
54. Li M, Zhu W, Gao L, Lu L. Changes in extracellular polysaccharide content and morphology of *Microcystis aeruginosa* at different specific growth rates. *J Appl Phycol*. 2013;25:1023–30. <http://dx.doi.org/10.1007/s10811-012-9937-7>.
55. Kehr JC, Zilliges Y, Springer A, Disney MD, Ratner DD, Bouchier C, et al. A mannan binding lectin is involved in cell-cell attachment in a toxic strain of *Microcystis aeruginosa*. *Mol Microbiol*. 2006;59:893–906. PubMed <http://dx.doi.org/10.1111/j.1365-2958.2005.05001.x>.
56. Hayes PK, Lazarus CM, Bees A, Walker JE, Walsby AE. The protein encoded by *gvpC* is a minor component of gas vesicles isolated from the cyanobacteria *Anabaena flos-aquae* and *Microcystis* sp. *Mol Microbiol*. 1988;2:545–52. PubMed.
57. Mlouka A, Comte K, Castets A-M, Bouchier C, Tandeau de Marsac N. The Gas Vesicle Gene Cluster from *Microcystis aeruginosa* and DNA rearrangements that lead to loss of cell buoyancy. *J Bacteriol*. 2004;186:2355–65. PubMed.
58. Dunton PG, Walsby AE. The diameter and critical collapse pressure of gas vesicles in *Microcystis* are correlated with GvpCs of different length. *FEMS Microbiol Lett*. 2005;247:37–43. PubMed.
59. Hauman J. Is a Plasmid(s) Involved in the Toxicity of *Microcystis aeruginosa*? In: Carmichael WW, editor. *The Water Environment*. US; 1981. p. 97–102. http://dx.doi.org/10.1007/978-1-4613-3267-1_7.
60. Schwabe W, Weihe A, Börner T, Henning M, Kohl J-G. Plasmids in toxic and nontoxic strains of the cyanobacterium *Microcystis aeruginosa*. *Curr Microbiol*. 1988;17:133–7.
61. Bolch CJS, Blackburn SI, Jones GJ, Orr PT, Grewe PM. Plasmid content and distribution in the toxic cyanobacterial genus *Microcystis* Kützing ex Lemmermann (Cyanobacteria: Chroococcales). *Phycologia*. 1997;36:6–11. <http://dx.doi.org/10.2216/i0031-8884-36-1-6.1>.
62. Darling AC, Mau B, Blattner FR, Perna NT. Mauve: multiple alignment of conserved genomic sequence with rearrangements. *Genome Res*. 2004;14:1394–403. PubMed <http://dx.doi.org/10.1101/gr.2289704>.
63. Roberts RJ, Vincze T, Posfai J, Macelis D. REBASE-a database for DNA restriction and modification: enzymes, genes and genomes. *Nucleic Acids Res*. 2010;38:D234–6. PubMed <http://dx.doi.org/10.1093/nar/gku1046>.
64. Roberts RJ, Vincze T, Posfai J, Macelis D. REBASE-enzymes and genes for DNA restriction and modification. *Nucleic Acids Res*. 2007;35:D269–270. PubMed <http://dx.doi.org/10.1093/nar/gkl891>.
65. Bland C, Ramsey TL, Sabree F, Lowe M, Brown K, Kyripides NC, et al. CRISPR recognition tool (CRT): a tool for automatic detection of clustered regularly interspaced palindromic repeats. *BMC Bioinformatics*. 2007;8:209. PubMed <http://dx.doi.org/10.1186/1471-2105-8-209>.
66. Haft DH, Selengut J, Mongodin EF, Nelson KE. A guild of 45 CRISPR-associated (Cas) protein families and multiple CRISPR/Cas subtypes exist in prokaryotic genomes. *PLoS Comput Biol*. 2005;1:e60. PubMed <http://dx.doi.org/10.1371/journal.pcbi.0010060>.
67. Dobrindt U, Hochhut B, Hentschel U, Hacker J. Genomic islands in pathogenic and environmental microorganisms. *Nat Rev Microbiol*. 2004;2:414–24. PubMed <http://dx.doi.org/10.1038/nrmicro884>.
68. Langille MG, Brinkman FS. IslandViewer: an integrated interface for computational identification and visualization of genomic islands. *Bioinformatics*. 2009;25:664–5. PubMed <http://dx.doi.org/10.1093/bioinformatics/btp030>.
69. Field D, Garrity G, Gray T, Morrison N, Selengut J, Sterk P, et al. The minimum information about a genome sequence (MIGS) specification. *Nat Biotechnol*. 2008;26:541–7. PubMed <http://dx.doi.org/10.1038/nbt1360>.
70. Woese CR, Kandler O, Wheelis ML. Towards a natural system of organisms: proposal for the domains Archaea, Bacteria, and Eucarya. *Proc Natl Acad Sci U S A*. 1990;87:4576–9. PubMed <http://dx.doi.org/10.1073/pnas.87.12.4576>.
71. McNeill J, Barrie FR, Burdet HM, Demoulin V, Hawksworth DL, Marhold K, et al. International Code of Botanical Nomenclature. Königstein: A.R.G. Gantner; 2006. p. 1.
72. Woese CR, Stackebrandt E, Macke TJ, Fox GE. A phylogenetic definition of the major eubacterial taxa. *Syst Appl Microbiol*. 1985;6:143–51. PubMed [http://dx.doi.org/10.1016/S0723-2020\(85\)80047-3](http://dx.doi.org/10.1016/S0723-2020(85)80047-3).
73. Stanier RY, Kunisawa R, Mandel M, Cohen-Bazire G. Purification and properties of unicellular blue-green algae (order Chroococcales). *Bacteriol Rev*. 1971;35:171–205. PubMed.
74. Komárek J, Anagnostidis K. Cyanoprokaryota 1. Teil: Chroococcales. In: Ettl H, Gartner G, Heynig H, Mollenhauer D, editors. *Süßwasserflora von Mitteleuropa*. 1999;p.164-169.
75. Geitler L. Cyanophyceae. In: Rabenhorst L, von Deutschland KF, editors. *Österreich und der Schweiz Leipzig: Akademische Verlagsgesellschaft*, vol. 14. 1932. p. 130–59.
76. Caumette P, Brochier-Armanet C, Normand P. Taxonomy and Phylogeny of Prokaryotes. In: Bertrand J-C, Caumette P, Lebaron P, Matheron R, Normand P, Sime-Ngando T, editors. *Environmental Microbiology: Fundamentals and Applications*. Netherlands: Springer; 2015. p. 177–9. http://dx.doi.org/10.1007/978-94-017-9118-2_6.
77. Via-Ordorika L, Fastner J, Kurmayer R, Hisbergues M, Dittmann E, Komarek J, et al. Distribution of microcystin-producing and non-microcystin-producing *Microcystis* sp. in European freshwater bodies: detection of microcystins and microcystin genes in individual colonies. *Syst Appl Microbiol*. 2004;27:592–602. PubMed <http://dx.doi.org/10.1078/0723202041748163>.
78. Ashburner M, Ball CA, Blake JA, Botstein D, Butler H, Cherry JM, et al. Gene ontology: tool for the unification of biology. *The Gene Ontology Consortium*. *Nature genetics*. 2000; 25:25–29. PubMed. <http://dx.doi.org/10.1038/75556>.