Responses of Cyanobacterial Aggregate Microbial Communities to Algal Blooms

Congmin Zhu
  Tsinghua University   https://orcid.org/0000-0002-0472-9128

Junyi Zhang
  Southeast University

Xin Wang
  Peking Union Medical College Hospital

Yuqing Yang
  Sogou Inc

Ning Chen
  Tsinghua University

Zuhong Lu
  Southeast University

Qinyu Ge
  Southeast University

Rui Jiang
  Tsinghua University

Xuegong Zhang
  Tsinghua University

Yunfeng Yang
  Tsinghua University

Ting Chen (✉ tingchen@tsinghua.edu.cn)
  Tsinghua University   https://orcid.org/0000-0002-3228-9166

Abstract

Background

Freshwater lakes are threatened by harmful cyanobacterial blooms; whose basic unit is Cyanobacterial Aggregate (CA). Community variations of CA-attached bacteria are substantial during different blooming stages. However, little is known about their transcriptional and metabolic variations. Most bacterial genomes in CA were not constructed in existing database, which limits our understanding of the bacterial variations as responses to cyanobacterial blooms.

Results

In this longitudinal study, 16 CA samples were collected from Lake Taihu, one of the largest freshwater lakes in China, from April of 2015 to February of 2016. By sequencing the V4 region of 16S rRNA genes, full metagenomes (MG) and metatranscriptomes (MT), we generated 424 Mb of 16S rRNA gene data, 122 Gb of high-quality MG data and 160 Gb of high-quality MT data. We analyzed the taxonomic, functional and transcriptional variations of microbes in CAs along three blooming stages, and constructed metagenome-assembled genomes (MAGs) by binning analysis. First, 55 OTUs, 456 genes and 37 transcripts mainly associated with pathways of transporters, photosystem and energy metabolism showed significantly different abundance among the three stages. Second, 161 high-quality MAGs in CAs were achieved, with 19 of which significantly shifted in relative abundance among three stages. The most abundant MAGs have gene capacities to synthesize flagella and divers of transporters, and participate in metabolic pathways of nitrogen, phosphorus and sulfur. Finally, 22 high-quality cyanobacterial MAGs were constructed and can be divided into four functional clusters, which showed significant differences on the energy pathways, transporters and prokaryotic defense system.

Conclusion

Overall, these results demonstrated the taxonomic, functional and transcriptional variations of microbes in CAs among three different blooming stages. Genome construction and metabolic analysis of cyanobacteria and their attached bacteria suggested that the material exchange and signal transmission do, indeed, exist among them. Our understanding of the underlying molecular pathways for cyanobacterial blooms could potentially lead to the control of blooms by interventional strategies to disrupt the expression of critical microbes.

Introduction

Eutrophication has threatened freshwater lakes worldwide, leading to Cyanobacterial Harmful Algal Blooms (CyanoHABs) [1, 2]. Lake Taihu, which is the third-largest freshwater lake in China, plays vital services in agricultural irrigation, fisheries and water supply to a population of 2 million citizens [3]. However, CyanoHABs has turned Lake Taihu into a toxic nightmare since the early 1990s by producing toxins and causing hypoxia in the water, which dramatically increased the cost for water treatment and reduced the drinking water supplies. Accelerated by exceeding nutrient loadings, recent years have witnessed that the blooming time expanded from summer to any season of the year [4, 5]. Therefore, it is imperative to elucidate the underlying mechanism for effective treatment strategies.

Cyanobacterial Aggregate (CA) is the basic unit wherein cyanobacteria and bacteria coexist in close contact, which is major form of cyanobacteria during blooms. As autotrophic organisms, cyanobacteria can provide food for their attached microbes by excreting rich extracellular organic matters. In turn, attached bacteria may supply nitrogen, phosphorous, sulfur and trace elements to the cyanobacteria [6, 7]. Metagenomics studies focused on CA-attached bacteria have given us important insights into the bacterial community structure in CA with reporting their seasonal variation and spatial distribution [8, 9]. Alternate succession of two cyanobacteria genera (*Microcystis* and *Dolichospermum*) was found and the attached bacteria in CA also changed along the cyanobacterial succession along different blooming stages [10]. However, it is unknown whether the genes and transcripts expressed by these bacteria will also change significantly at different stages. Additionally, mutualistic relationships between cyanobacteria and many attached bacteria have been proposed [11-14]. However, it failed to explain the phenomenon from the perspective of microbe genes and transcripts, because of the lack of reference genome which are critical for uncovering the cellular and molecular mechanisms.

In this study, we focused on the CA communities and performed multi-omics of microbiomes during different cyanobacterial blooming stages. To this end, we collected 16 CA samples from Lake Taihu in different cyanobacterial blooming stages and then sequenced 16S rRNA genes, metagenomes and metatranscriptomes for multi-omics analysis. In this report, we were able to (1) reveal the variation of taxonomic, functional and transcriptional compositions of CAs along different blooming stages, (2) construct the draft genomes of microbes in CAs to reveal the functional links between cyanobacteria and their attached bacteria, and (3) compare the cyanobacterial strains and position their ecological roles in the blooming.

Materials And Methods

Sample collection and remote sensing data

From March 4, 2015 to January 27, 2016, a total 16 CAs samples covering the yearly cycle of a CyanoHAB were collected from Lake Taihu. These samples were collected using weekly time-series when CyanoHAB occurred from four sampling sites shown in Table 1. To extract CAs samples, water at 0.5m depth was collected by using 5 L Schindler sampler and then filtered by 40 μm pore size Cell Strainer (BD Falcon) and processed according to the previous study [15]. In brief, the aggregates were collected by micropipetting into tubes after the samples were kept motionless until most CAs had floated to the surface. Then, we adopted the vortex and washing method and repeated it multiple times to remove as much free-living bacteria as possible while maintaining cellular contents of cyanobacterial aggregates. Finally, about 10 ml of wet CAs were transferred to sterile Eppendorf tubes and stored at -80°C for DNA extraction.

We used remote sensing method to monitor the CyanoHABs in Lake Taihu to provide real-time monitoring of the whole blooming area. The remote sensing image data of Lake Taihu came from MODIS sensor of NASA and were analyzed by ENVI (Environment for Visualizing Images, version 4.2) [16]. We calculated the covering area of cyanobacteria in Lake Taihu and summarized the pattern of cyanobacteria bloom in a year. Besides, the total nitrogen (TN) of related CAs samples were measured according to the standard methods of water analysis [17], and water temperature (WT) and Chemical oxygen demand (COD), which were measured with a YSI 6600 Multi-Parameter Water Quality Sonde [18].

DNA extraction and mRNA extraction

To extract DNA from CAs, 200 mg of CAs were first lysed by guanidine thiocyanate and N-lauroyl sarcosine (SIGMA) and then incubated in 70°C for 1 hour, followed by vortexing with beads and centrifugation. The supernatant was transferred into a new tube with 500 μl of extraction buffer (TENP mixture, including ddH2O, Tris-HCl, EDTA, NaC, PVPP and water), vortexed and centrifuged before 500 μl supernatant were collected. This step was repeated three times. All supernatants were then centrifuged at high speed, and the resulting supernatant was precipitated with isopropanol overnight at 4°C. The mixture was then centrifuged at high speed, and the supernatant was discarded. The pellet was chilled on ice for several hours after the addition of phosphate buffer and potassium acetate. Two volumes of ethyl alcohol and 0.1 volume of sodium acetate were added to the mixture, which was then stored at -20°C for several hours. The mixture was centrifuged at high speed, and the DNA deposit was washed twice with 70% ethanol. DNA was dissolved and preserved in TE buffer (AM9849).

To extract mRNA from CAs, the collected samples were immediately placed in TRIzol Reagent (Ambion, Foster City, CA, USA) (adding 1 ml TRIzol reagent per 50–100 mg of the sample) to prevent RNA degradation and were stored at - 80 °C until extraction. Total RNA was extracted according to the manufacturer protocol, followed by NanoDrop 1000 Spectrophotometer (NanoDrop Technologies, Wilmington, DE, USA) and 1% agarose gel electrophoresis (AGE) [19] to examine the quality and quantity. Considering rRNA account for more than 80% of the content of total RNA in biological cell, the Ribo-Zero™ rRNA Removal Kits (Illumina) were used to remove most of rRNA and get nearly neat mRNA. The extracted mRNA was dissolved in RNase-free water (Sigma, MO).

Library construction and sequencing

We took culture-independent multi-omics approaches: targeted 16S rRNA gene sequencing (16S) to determine the microbial community composition; total metagenome sequencing (MG) to determine the phylogenetic and functional genes, and total metatranscriptome sequencing (MT) to determine activity of genes. We also compared the different molecular approaches among each other. For 16S, we amplified the V4 region (250 bp) to assess the taxonomic composition of the CAs. PCR conditions and primer sequences were performed as previously described [10]. The purified amplicons were sequenced on an Illumina MiSeq PE250 sequencer at the Beijing Genomics Institute, Shenzhen, China (BGI). The 16S data have been submitted to the Sequence Read Archive (BioProject accession number PRJNA664292). For MG, the metagenome DNA libraries were constructed with 2 μg of DNA genomes, according to the Illumina TruSeq DNA Sample Prep v2 Guide, with an average of 350 bp insert size. The quality of all libraries was evaluated using an Agilent bioanalyzer with a DNA LabChip 1000 kit. All qualified libraries were loaded to Illumina HiSeq2500 for sequencing. The MG reads were deposited to the Sequence Read Archive (BioProject accession number PRJNA664299). For MT, the metatranscriptome mRNA libraries were constructed with ScriptSeq™ v2 RNA-Seq Library Preparation Kit (Illumina), including cDNA synthesis, end repair, dA-tailing, and adaptor ligation. The libraries were sequentially linearly amplified and qualitied with Qubit® 2.0 Fluorometer with the Qubit® HSDNA BR assay kit (Invitrogen™, Eugene, OR, USA) and Agilent bioanalyzer (Agilent Technologies, Santa Clara, CA, USA), respectively. All qualified libraries were loaded to Illumina HiSeq2500 to be sequenced with 2 × 125 bp pair-end technology. The MT reads were deposited to the Sequence Read Archive (BioProject accession number PRJNA664620).

16S rRNA gene sequence data processing

The raw data of 16S rRNA genes of CA samples were analyzed using Quantitative Insights into Microbial Ecology (QIIME) [20]. In brief, sequences with quality scores lower than 20 and length shorter than 200 were filtered out. Then Operational Taxonomic Units (OTUs) were identified using an open-reference OTU picking protocol at a 97% similarity cutoff. We mapped the OTUs to the Greengenes database (V201305) [21] to identify their taxonomic origins, and removed OTUs without alignment results. We estimated raw abundance of OTUs in each sample by counting the read numbers and conducted diversity analyses. After obtaining the OTU tables and phylogenetic trees, we calculated microbial richness estimators (Observed OTUs, PD whole tree and Chao1), evenness estimators (Equitability), and diversity estimators (Simpson Index) by using the R statistical calculation software package 'vegan' [22, 23].. To generate rarefaction curves, we randomly selected a fixed number of sequences from each dataset, using the lowest read number in our samples (~25k sequences) to rarify the OTU abundance data.

Metagenomic data processing

Illumina raw reads were filtered with the following constraints: (1) reads with more than 2 ambiguous N bases were removed; (2) reads with less than 80% of high-quality bases (Phred score ≥20) were removed; (3) 3'-ends of reads were trimmed to the first high-quality base. Then, filtered metagenomic reads were assembled by Megahit (version 1.0.5) [24] into contigs in a time- and cost-efficient way, using the following parameters: −min-contig-len = 150, --k-min = 27, --k-max = 123, --k-step = 8 and −min-count = 1. All assembled contigs were submitted to MetaProdigal (version 2.6.3) [25] for gene calling using the default parameters. We aligned all reads to genes by Bowtie2 and calculated the gene coverage using bedtools (version 2.26) [26]. We mapped the predicted genes to NCBI bacteria, archaebacterial and virus non-redundant genome databases by Diamond [27]. The alignment result was then submitted to Megan (version 6) [28] to estimate the taxonomic and functional compositions with weighted LCA algorithm. The functional analysis was conducted by mapping genes to Kyoto Encyclopedia of Genes and Genomes (KEGG) [29] and SEED [30]. Furthermore, Phosphate, nitrogen and sulfur were thought to be the key nutrients responsible for the cyanobacteria proliferation and blooming. Microcystins, which are the most common cyan-toxins in Lake Taihu, have strong hepatotoxicity that can severely damage mammalian liver cells. They may play a pivotal role in development of Microcystis colonies [87, 89]. We used Wilcoxon rank-sum test to select the related genes in different abundance (relative abundance ratios > 2).

To identify novel genomes from metagenomic sequencing data, contigs longer than 1 kbp were used with CONCOCT (version 0.4.1) [31] to recover microbial population genomes based on sequence compositions and coverages across multiple samples. Before binning, Bowtie2 [26] was used to align short-read sequences to contigs (options: -very-fast) and SAMtools [32] was used to sort and convert SAM files to BAM format. Sorted BAM files were then used to calculate the coverage of each contig in each sample metagenome. Then, we applied CheckM (version 1.0.6) [33] to estimate the completeness and contamination for each bin. The bins with high completeness (>70%) and few contaminants (<10%) were retained as MAGS. To get the nonredundant consolidation, the dRep [34] dereplication workflow was used with options 'dereplicate_wf -p 16 -comp 70 -con 10 -str 100 -strW 0'. Thus, in prefiltering, only bins assessed by CheckM as having both completeness ≥70% and contamination <10% were retained for pairwise comparison. Bin scores were given as completeness-5×contamination+0.5×log(N50), and only the highest scoring MAGs from each secondary cluster was retained in the dereplicated set. For the nonredundant MAGs, we inferred the phylogenetic relationship using bacterial core gene set that is defined by up-to-date bacterial genome database by UBCG [35] and drawn the phylogenetic tree using Evolview [36]. We calculated their abundance by bowtie2 and samtools. Protein-encoding genes from MAG contig sequences were predicted with MetaProdigal (version 2.6.3) [25] and the resulting nucleotide sequences were searched against the NCBI bacterial, archaeal and viral non-redundant genome databases using DIAMON [27]. The functional analysis was conducted by mapping genes to Kyoto Encyclopedia of Genes and Genomes (KEGG) [29] and SEED [30].

Metatranscriptomic data processing

The quality control of the Metatranscriptomic Illumina raw reads was the same criteria as that of metagenome sequencing data. Then, the reads were mapped to rRNA and SILVA database by using blastn to identify those rRNA, followed by Perl to separate the rRNA from other RNAs. Then, filtered mRNA reads were assembled by Trinity [37, 38] using the default parameters. The gene prediction and functional analysis were performed with the same steps as Metagenomic data processing. Additionally, we estimated gene activity by calculating the ratio of gene transcript abundance over gene sequence abundance, annotated as MT/MG [39].

Microcystis and *Dolichospermum* with various strains are the dominant genera during cyanobacterial blooms in Lake Taihu. To recognize outbreak strains of *Microcystis aeruginosa* and *Dolichospermum flos-aquae*, PanPhLAn [40]was applied for strain-level microbial profiling for each sample using metagenomic data. The pan-genome of *Microcystis aeruginosa* was built by extracting all genes from four available whole reference genomes (*M. aeruginosa NIES-2481*: GCA_001704955.1, *M. aeruginosa NIES-2549*: GCA_000981785.1, *M. aeruginosa NIES-*

843: GCA_000010625.1 and *M. aeruginosa PCC-7806SL*: GCA_002095975.1) and merging them into gene family clusters by USEARCH with default threshold of >=95%. The pangenome of *Dolichospermum flos-aquae* was builtd from the available genome of *D. flos-aquae CHAB1629*. Then, the strain-specific gene repertoires within a metagenomics sample were identified by PanPhlAn based on gene family co-abundance, with the assumption that single-copy genes from the same genome should have comparable sequencing coverage within the sample.

Comparative analysis between different blooming stages

Sample grouping

Using real-time remote sensing image data of Lake Taihu, we calculated the maximum aggregation area of the cyanobacterial blooms and frequency of cyanobacterial blooms per month. Missing data occurred because of that remote sensing data was not available when sampling in cloudy days. We replaced it with the data of the closest day from the sampling time. According to the remote sensing data, the blooming area in Lake Taihu dramatically increased in June and remained relatively steady until January to start decreasing (Table 1). Therefore, the blooming cycle is divided into three stages: Early blooming stage (Early-stage) from April to May, Middle blooming stage (Mid-stage) from June to December, and Late blooming stage (Late-stage) from January to February (Table 1).

Comparative analyses

Considering the bias of sample number between different groups, we randomly under-sampled the samples of Mid-stage to three samples. To use more samples as possible, the under-sampling were repeated three times and the results were averaged. Differences in the relative abundance of OTUs, genes and transcripts among three stages were analyzed using STAMP (V.2) [41]. ANOVA test and Welch's *t*-test were respectively used to compare among three groups and between two groups. FDR was implemented for multiple testing correction and FDR *p*-vlaue <0.05 was used to indicate statistical significance for all statistic tests. The taxa of differentially abundant OTUs were annotated with GreenGene database, while the functions of differentially abundant genes and transcripts were annotated with KEGG. Alpha-diversity (Shannon index) values were computed and compared for the three stages.

Results

Data description and blooming stage division

We collected 16 CAs from the northern bays of Lake Taihu from April 2015 to February 2016, in which cyanobacteria showed a full bloom cycle (Table 1). All samples were used for 16S rRNA gene sequencing to detect bacterial community composition, metagenome sequencing (MG) to detect functional potentials and metatranscriptome sequencing (MT) to detect community gene expression. A total of 424 Mb of 16S rRNA gene amplicon sequence data were generated from Illumina MiSeq sequencing, and 122 Gb of high-quality MG sequence data and 160 Gb of high-quality MT sequence data were generated from Illumina HiSeq sequencing, which are among the biggest dataset obtained for any CAs so far. Using real-time remote sensing image data of Lake Taihu, we calculated the total area of cyanobacterial blooming. The blooming area was very small (~4.5 km^2) in April and May, and expanded dramatically (~129.1 km^2) in June, and continued to decline (~1.70 km^2) in January. Therefore, we divided the blooming cycle into three stages (Early-stage: pre-blooming stage, Mid-stage: peak blooming stage and Late-stage: post-blooming stage). These 16 CA samples spanned the three blooming stages: three samples in Early-stage, 10 samples in Mid-stage and three samples in Late-stage (Table 1). Considering the imbalance of sample size in three groups, we corrected the sample size in all subsequent analyses.

Dynamics of microbial taxonomic structure in CAs

Bacteria dominated the CA community (MG 97.561.89%, MT 86.496.77%), with small fractions of Archaea (MG 0.190.07%, MT 0.380.13%) and Eukaryota (MG 2.251.84%, MT 13.136.68%). The most abundant bacteria found in the Early-stage samples on the basis of 16S were members of *Cyanobacteria* (38.85.1% s.d.), *Bacteroidetes* (30.010.1%) and *Proteobacteria* (26.54.8%) (Fig. 1A). The most abundant bacteria found in the Mid-stage samples were members of *Proteobacteria* (36.16.2%), *Cyanobacteria* (35.88.5%) and *Bacteroidetes* (20.0%). The most abundant bacteria in the Late-stage samples were members of *Cyanobacteria* (68.23.7%), *Proteobacteria* (17.84.6%) and *Bacteroidetes* (12.52.2%). Similar distributions of phyla were observed in 16S, MG and MT data sets (correlation coefficients of r>0.65 for 16S and MG, r>0.85 for MG and MT, and r>0.90 for 16S and MT), suggesting that the data sets were of excellent quality.

Different stages had unique sets of microbes, especially for Mid-stage (Fig. 1B). When comparing the taxonomic composition among three stages, 50 OTUs (22 Bacteroidetes,10 Cyanobacteria, 15 Proteobacteria and 3 Unassigned) showed significantly different abundance among three stages via ANOVA test (Fig. 2A), 16 of which belong to the family *Cytophagaceae* and four of which belong to the cyanobacterial genus *Dolichospermum*. According to UPGMA clustering results of the OTUs abundance, 16 samples could also be divided into three significantly different groups (Adonis, R^2 = 0.71, p=0.001). Remarkably, the three groups correspond to three blooming stages. When comparing the taxonomic composition between pairwise stages, four OTUs showed different abundance between Early-stage and Mid-stage; seven between Early-stage and Late-stage; one between Mid-stage and Late-stage (Fig. S1). It is noteworthy that the cyanobacterial genus *Microcystis* was detected enriched in Mid-stage than Early-stage (Fig S1A), which is consistent with previous studies showing that *Microcystis* was the dominant genus during heavy cyanobacterial blooming in summer and autumn [36, 86, 87].

Different alpha diversity was detected in three stages based on OTU abundance table, with Early-stage exhibiting the highest alpha diversity (Fig. 1C). Early-stage and Mid-stage were similar in 16S rRNA operon copy number and total nitrogen. However, it is interesting that the alpha diversity was the lowest in the Late-stage, but 16S rRNA operon copy number and total nitrogen were the highest, consistent with the resource-copy number theory ("more resource, more copy number") [42] and the resource-ratio theory ("more resource, less diversity") [43] .

Dynamics of microbial functional structure in CAs

The CAs of three blooming stages shared more genes and transcripts compared to OTUs (Fig. 1B). The CAs in Mid-stage exhibited more genes and transcripts than those in the other two stages. While, the CAs in Early-stage and Late-stage shared least amount of genes and transcripts, consistent with the conclusion from OTUs analysis. When annotated the genes and transcripts with KEGG database, the CAs in Mid-stage and Late-stage shared more transcripts that mainly include nitrogen metabolisms related genes, such as denitrification and ammonification. (Fig. 1D). A total of 456 genes mainly related to enzymes (244 genes), ribosome (46 genes) and transporters (42 genes) showed significantly different abundance among three stages (ANOVA test, Fig. 2B). Meanwhile, 37 transcripts mainly related to enzymes (15 genes), ribosome (9 genes) and photosynthesis proteins (5 genes) had significantly different abundance among three stages, of which 20 transcripts (Fig. 2C, highlight with light yellow) were annotated to the same KOs of the significant genes. Those 20 transcripts were mainly included in the pathway of photosynthesis system and subunit ribosomal protein. It is noteworthy that most of the genes (436 of 456) with significant difference were not differently expressed. Six SEED subsystems had significantly different abundance among three stages (Fig. S2A). Photosynthesis related subsystems, such as electron transport and photophosphorylation, NAD and NADP, light-harvesting complexes and alkylphosphonate utilization, are significantly enriched in Mid-stage compared to Early-stage (Fig. S2B). Protein secretion system is significantly enriched in Early-stage; meanwhile ABC transporter is significantly enriched in Early-stage compared to Mid-stage (Fig. S2B).

Phosphate, nitrogen and sulfur were closely related to cyanobacterial bloom; which give rise to methane and algal toxin production. Therefore, we focused on the genes on the related pathways. The pentose phosphate pathway, sulfur metabolism and methane metabolism appeared in all three stages from both MG and MT data, with the Mid-stage showing the most abundant genes on all these pathways (Fig. 3). Dissimilatory reduction of nitrate to ammonium, nitrogen fixation and denitrification were detected and nitrogen fixation only appeared in Late-stage. Microcystins are a group of at least 80 chemical variants and synthesized non-ribosomally by protein products of gene cluster *mcy*. The *mcy*s were detected in all three stages from both MG and MT data and most abundant in Mid-stage, followed by Late-stage and Early-stage.

When using the MT/MG ratio to estimate taxonomic activity at the genus level, the average MT/MG ratios were the highest for *Basidiomycota* (MT/MG=4.1), *Ascomycota* (MT/MG=3.5) and *Cyanobacteria* (MT/MG=1.9), suggesting that *Cyanobacteria*, in addition to the first two fungi, were acclimated to be the most active (Fig. 1E). However, considering the relative abundance of eukaryote resulted from metagenomic data is 2.2%, we focused on the bacteria. We calculated the MT/MG ratio for each genus of bacteria. As shown in Table 2, the genera showing highest activity in three blooming stages were different. The three most active genera for Early-stage are *Pseudomonas* (MT/MG=3.8), *Dolichospermum* (MT/MG=3.2) and *Pseudanabaena* (MT/MG=3.0). *Dolichospermum* is the dominant cyanobacterial genus in the Early-stage resulted from 16S rRNA analysis and were acclimated to be active (Fig. 1E). In Mid-stage, the three most active genera are *Arcobacter* (MT/MG=12.5), *Dechloromonas* (MT/MG=5.7) and *Clostridium* (MT/MG=4.5). All the three genera are cyanobacterial attached bacteria and are more active than cyanobacteria, suggesting their significant roles in heavy water bloom. The. In Late-stage, *Dolichospermum* (MT/MG=4.4), *Aphanizomenon* (MT/MG=4.0) and *Pseudomonas* (MT/MG=4.0) are the highest active genera. Different from the Mid-stage, the nitrogen fixation cyanobacteria genus, *Dolichospermum*, returned to be the dominant cyanobacterial genus in Late-stage and shows highest MT/MG ratio. Another nitrogen fixation cyanobacterial genus, *Aphanizomenon* is the second most active genus. Several genes involved in nitrogen fixation were also detected in the MG and MT data sets.

Metagenome-assembled genomes (MAGs)

We generated 233 metagenome-assembled genomes (MAGs) with ≥60% completeness and <10% contamination, 161 of which were high quality draft genomes with ≥70% completeness and <5% contamination [44]. De-replication of highly similar MAGs based on FastANI values of ≥95% resulted in the consolidation of 78 non-redundant MAGs and their phylogenetic tree (Fig. 4). The tree is dominated by large numbers of genomes from Proteobacteria (44 MAGs, mostly Alphaproteobacteria) and Bacteroidetes (15 MAGs, mostly Cytophagia) phyla, and also contains several genomes from the phyla of Gemmatimonadetes (2 MAGs), Spirochaetes (2 MAGs), Cyanobacteria (2 MAGs), Acidobacteria (1 MAG) and 12 novel Bacterial genomes. Their completeness, contamination, length, N50 and taxonomic annotation were shown in Table 3.

The relative abundance of 78 non-redundant MAGs in all samples is provided in Table S1. Using a cut-off of 1×coverage, most MAGs (72) were present in more than one samples and 26 MAGs were present in more than eight samples. Four MAGs were present in more than 14 samples, which were all members of the Alphaproteobacteria. Nineteen MAGs were shifted in relative abundance among three stages (ANOVA test, $P < 0.10$) (Fig. 4) and their relative abundance were shown in Fig. S3A. Two MAGs annotated to Cytophagales order and Caulobacteraceae family showed higher abundance in Early-stage than Mid-stage (Welch's t-test, $P < 0.05$), while two MAGs belonging to Acetobacteraceae family and Alphaproteobacteria class showed higher abundance in Mid-stage than Early-stage (Fig. S3B). Ten MAGs showed significantly different abundance between Early-stage and Mid-stage, only one of which (150608.59.fa annotated to Cyanobacteria phylum) had higher abundance in Late-stage. When comparing between Mid-stage and Late-stage, 15 MAGs were more abundant in Mid-stage, except one MAGs (150608.59.fa annotated to Cyanobacteria phylum) that was more abundant in Late-stage. The cyanobacterial MAG 150608.59.fa has closest sequence similarity to *Dolichospermum flos-aquae.CHAB-1629* (Genome-to-Genome Distance Calculator [45] DNA−DNA hybridization (DDH) 18.4%2.4). *Dolichospermum* was detected to be the dominated cyanobacterial genus in the Late-stage, with 150608.59.fa showing the highest abundance in Late-stage. The other non-redundant cyanobacterial MAGs (150402.41.fa with the closest sequence similarity to *Microcystis aeruginosa.NIES-88*, DDH 12.9%2.7), was detected enriched in Mid-stage than Early-stage.

Based on the average relative abundance, the most abundant MAG was 151222.71.fa annotated to Betaproteobacteria. We constructed its metabolic pathways (Fig. 5) and identified rich transporters of ammoniu, phosphate, amino acid, peptide and sugar, and several metallic elements such as iron, nickel, cobalt and magnesium. It has the flagellar biosynthesis pathway to synthesize flagellar to improve movement in viscous environments. It also has the nitrogen metabolism transferring nitrate to ammonia, the sulfur metabolism transferring thiosulfate to sulfate, the fatty acid metabolism transferring acetyl-coa to plamitic acid and the oxidative phosphorylation pathway releasing energy supplies ADP and inorganic phosphoric acid to synthesize ATP through respiratory chain.

Cyanobacterial strains

Twenty two high-quality cyanobacterial MAGs were constructed from binning analysis, which were clustered into four clusters based on their gene function (Fig. 6). Cluster 1 and Cluster 2 only contain one MAG. Cluster 3 and Cluster 4 respectively contain nine and 11 MAGs. Their phylogenetic tree, alongside 21 public genomes of two dominant cyanobacterial genera (*Microcystis* and *Dolichospermum*) in Lake Taihu from NCBI, was constructed based on bacterial core gene set (Fig. S4). As expected, these 22 cyanobacterial MAGs could also be divided into four evolutionary divergence in the phylogenetic tree. Both Cluster 1 and Cluster 2 are similar to *Microcystis aeruginosa*. Cluster 3 is close to *Dolichospermum flos-aquae* and Cluster 4 to *Pseudanabaena yagii GIHE-NHR1*. There were significant differences in metabolism among the four clusters, especially on the energy pathways of nitrogen metabolism, sulfur metabolism and photosynthesis (Fig. 6). Cluster 1 and Cluster 2 have more energy metabolism pathways than Cluster 3 and Cluster 4, with the modules of M00145 (NAD(P)H:quinone oxidoreductase, chloroplasts and cyanobacteria), M00163 (Photosystem I), M00161 (Photosystem II) and M00616 (Sulfate-sulfur assimilation). However, Cluster 1 has extra sulfur metabolic modules (M00596: Dissimilatory sulfate reduction, M00176: Assimilatory sulfate reduction) and Cluster 2 has extra nitrogen metabolic modules (M00531: Assimilatory nitrate reduction, M00615: Nitrate assimilation). The energy pathways of Cluster 3 mainly include oxidative phosphorylation and photosynthesis (Photosystem I), while Cluster 4 mainly include Photosynthesis (Photosystem II), nitrogen and sulfur metabolism. All of the four clusters have prokaryotic defense system, especially Cluster 4, which has three CRISP-associated proteins Csm1, Cmr3 and Cmr4 (Fig. S5). Those cyanobacterial MAGs also have dramatically different transporters. Cluster 1 can transport phosphate, ammonium, peptide/nickel, biopolymer, ferrous iron, lipopolysaccharide and urea. Cluster 4 can transport high-affinity iron, Na+, lipopolysaccharide, simple sugar, putative Mg2+, NitT/TauT, biopolymer and urea. Cluster 2 can transport branched-chain amino acid, polar amino acid, neutral amino acid, cobalt/peptide/nickel, glutamate, chromate, Ca, Fe-S cluster assembly and zinc. Cluster 3 can transport vitamin B_{12}, MFS, arginine, lysine, histidine and glutatmine.

Microcystis aeruginosa strains varied by their capabilities on microcystins biosynthesis [46].We obtained the major strain for each sample by constructing strains generation for each metagenomic CA sample, and checked its toxicity by detecting *mcy* gene family (Fig. S6A). First, 10,568 gene families (similarity > 95%) were clustered from four reference genomes of *Microcystisaeruginosa*. Then, 12 *Microcystis aeruginosa* strains were obtained from 12 CA samples and they were clustered into two major clusters respectively similar to *M. aeruginosa NIES-843* and *M. aeruginosa PCC-9806*, which were dramatically different from the two very similar genomes of *M. aeruginosa NIES-2481*

and *M. aeruginosa NIES-2549*. Two strain clusters showed different microcystins biosynthesis characteristics. The cluster similar to *M. aeruginosaNIES-843* contained *mcy* gene family and the other cluster similar to *M. aeruginosa PCC-9806* was without *mcy* gene family. Additionally, different strains were composed of different gene families, indicating the high functional diversity of *Microcystis aeruginosa* strains in Lake Taihu. For the other dominant cyanobacterial genus *Dolichospermum flos-aquae*, 4,326 gene families were clustered from one available reference genome (*D. flos-aquae CHAB 1629*). Compared to *Microcystis aeruginosa*, *Dolichospermum flos-aquae* showed much lower strain diversity in Lake Taihu with almost same gene family combination to the reference genome of *D. flos-aquae CHAB 1629* (Fig. S6B). Seven major strains were constructed from seven CA samples, of which three samples were from Early-stage and three samples from Late-stage, consistent with the phenomenon that *Dolichospermum* showed significant higher abundance in Early-stage and Late-stage.

Discussion

The real-time remote sensing image data of Lake Taihu showed the significantly different cyanobacterial blooming in different month. The seasonal dynamics of microbial communities is apparent, including the succession of dominant cyanobacterial genera [47-49] and the variation of attached bacteria [47, 50, 51]. Four OTUs of cyanobacterial genus *Dolichospermum* were enriched in Early-stage and Late-stage, while the cyanobacterial genus *Microcystis* was more enriched in Mid-stage than Early-stage, demonstrating the alternate succession of dominant cyanobacterial genera between *Microcystis* and *Dolichospermum*. There are 55 OTUs as the attached bacteria, also showed significantly different abundance among three stages and 16 of them were belong to the family *Cytophagaceae* (Fig. 2A). All known members of *Cytophagaceae* are heterotrophic and many members digest macromolecules such as polysaccharides or proteins [52], which are secreted by cyanobacterial cells. Some *Cytophagaceae* are diazotrophs and some are proficient at digestion of insoluble cellulose in CA, which means that their abundances will change with the nitrogen content and cyanobacterial composition [53, 54]. Therefore, it might be used as the signal bacteria for the large-scale outbreak of cyanobacteria bloom.

Although the abundance of many genes varies significantly along the dynamics of microbial composition, only some transcripts changed significantly, especially those related to photosystem and subunit ribosomal protein. Photosystems is related to longer day time, and the subunit ribosomal proteins are related to ribosomes, which means more translations, more proteins being produced, more biomass, and faster growth. At the level of SEED subsystem, protein secretion system and ABC transporter were significantly enriched in Early-stage. These two subsystems are related to the protein exchange among microbes, which means that there is a frequent exchange of substances and signal transmission between microbes in the Early-stage. Photosynthesis related subsystems were significantly enriched in Mid-stage, which was most likely resulted from the mass propagation and growth of the cyanobacteria during heavy blooming in Mid-stage. The Mid-stage exhibited both more genes and transcripts than the other two stages, showing it has more functional diversity to adapt to the heavy cyanobacterial bloom. Considering that different cyanobacteria dominate different blooming stages, bacterial species and strains that carried out similar metabolic functions were likely to colonize similar algal taxa [6]. When comparing pairwise stages, the Early-stage and Late-stage shared fewest genes and transcripts, consistent with the conclusion drawn from taxonomy structure that community succession exists along cyanobacterial blooms [47, 55, 56].

When calculating the MT/MG ratios to estimate taxonomic activity at the genus level, the three most active genera for Early-stage are *Pseudomonas*, *Dolichospermum* and *Pseudanabaena*. *Dolichospermum*, as a nitrogen fixation cyanobacteria genus, is the dominant cyanobacterial genus in Early-stage. However, the genus with the highest gene activity is *Pseudomonas*, which can produce exopolysaccharides contributing to surface-colonizing biofilms, closely attached to cyanobacterial cells [57]. *Pseudomonas* is a versatile taxon capable of making full use of resources in cyanobacterial aggregates [58]. Additionally, the third most active genus *Pseudanabaena*, known to associate with water bloom, usually was covered by the gelatinous sheath of cyanobacteria because of its small cell size [59]. The Mid-stage exhibited much higher gene activity than the other two stages, with three most active genera *Arcobacter*, *Dechloromonas* and *Clostridium*. Some species of *Arcobacter*, which showed much higher activity than other members in cyanobacterial aggregates in Mid-stage, have genes related to nitrate reduction, nitrogen fixation and sulfide oxidation [60]. Through nitrogen and sulfide metabolism, *Arcobacter* interact with cyanobacteria and thus have a survival competitive advantage in cyanobacterial aggregates and consequently show high activity. *Dechloromonas*, a genus of *Betaproteobacteria*, is a dissimilatory (per)chlorate reducing bacterium [61]. *Clostridium*, as heterotrophic bacteria, can hydrolyze sugar and protein [62]. Exercise with the circumferential flagellum [63], *Clostridium* can maximally utilize the organic compound in the cyanobacterial aggregates to maintain high activity. In Late-stage, two enriched OTUs were annotated to the family *Comamonadaceae*, which are aerobic and harbor a remarkable phenotypic diversity that include anaerobic denitrifies, photoautotrophic and photoheterotrophic bacteria [64]. In Late-stage, *Dolichospermum*, *Aphanizomenon* and *Pseudomonas* are the highest active genera. Both *Dolichospermum* and *Aphanizomenon* are nitrogen fixation cyanobacteria genera [65]. A large proportion of fixed nitrogen may be released into the cyanobacterial aggregates by the two highest active genus, providing an important source of biologically-available nitrogen to the ecosystem and explaining the high measurements of total nitrogen in the Late-stage (Fig. 1C). Besides, *Aphanizomenon* become the second active genus may be because its ability to induce phosphate-limitation in other phytoplankton while also increasing phosphate availability to

itself through release of cylindrospermopsin [66]. For all stages, the most active genus and the most abundant genus basically were different, consistent with previous report about microbial activity [39].

The four clusters of cyanobacterial MAGs showed significant differences on the energy pathways and transporters. Cluster 1 and Cluster 2, similar to *M. aeruginosa*, have abundant energy metabolism pathways of nitrogen metabolism, sulfur metabolism and photosynthesis. The dynamic genotypic and phenotypic nature of *Microcystis* is evident in the high abundance of insertion sequences, transposable elements, restriction enzymes, and genes likely acquired through lateral transfer, which resulted varied strains [67]. Cluster 3, similar to *Dolichospermum*, only has oxidative phosphorylation and photosynthesis, but can transport vitamin B_{12}. It has been proposed that cyanobacteria obtain vitamins from synthesizing bacteria associated with them and that the heterotrophic bacteria were awarded with photosynthate in return [68]. This hypothesis is verified by our finding that cyanobacterial MAG can synthesize vitamin B_{12} transporter and protein secretor. Cluster 4, similar to *Pseudanabaena yagii GIHE-NHR1*, has three CRISPR-associated proteins Csm1, Cmr3 and Cmr4. Currently, at least 45 families of CRISPR-associated (Cas) proteins have been identified, including the CMR (Cas module RAMP (repeat-associated mysterious proteins)) and the CASCADE (CRISPR-associated complex for antiviral defense) complexes [69]. Cmr proteins were related to maturation, regulation of expression, Cmr complex formation or stabilization of CRISPR3 transcripts [70]. Our results verify that the photosynthetic cyanobacteria lack homologs to those Cas proteins commonly associated with the CASCADE complex in bacteria, but possess Cmr proteins instead [71].

Additionally, the most abundant MAG (151222.71.fa annotated to Betaproteobacteria) not only can synthesize flagella and mass of transporters, but also has the energy pathways of nitrogen, phosphorus and sulfur. During the cyanobacterial blooms, algae usually organized and formed viscous aggregates via secreting extracellular polysaccharides. Flagella can help the bacteria move better in such viscous environment. Rich transporters allow the bacteria to exchange nutrient and organic matter with cyanobacteria. It can also transport several metallic elements, such as iron, nickel, cobalt and magnesium, signifying the importance of trace metal elements on the growth and proliferation of cyanobacteria [72]. Additionally, it has the nitrogen metabolism transferring nitrate to ammonia, the sulfur metabolism transferring thiosulfate to sulfate, the fatty acid metabolism transferring acetyl-coa to plamitic acid and the oxidative phosphorylation pathway releasing energy supplies ADP and inorganic phosphoric acid to synthesize ATP through respiratory chain. This finding proven close bacterial associations with algae that are often highly evolved [73].

Conclusions

The taxonomic, functional and transcriptional variations of microbial community in cyanobacterial aggregates along different blooming stage shows the responses of microbes to algal blooms. The constructed high-quality MAGs are highly evolved with diverse of transporters and energy pathways, indicating the functional links by which to form mutualistic relationships between cyanobacteria and their closely attached bacteria. Cyanobacterial MAGs are clustered into four clusters with different functional genes on the energy pathways, transporters and prokaryotic defense system, which can explain the prevalent phenomenon of alternate succession of different cyanobacterial genera or species driven by environmental factors. Genome construction and metabolic analysis of cyanobacteria and their attached highly evolved bacteria suggested the material exchange and signal transmission among them, providing biological clues potentially leading to the control of blooms by interventional strategies to disrupt the expression of critical microbe.

Abbreviations

CAs: Cyanobacterial aggregates; rRNA: Ribosomal RNA; CyanoHABs: Cyanobacterial Harmful Algal Blooms; MG: Metagenome; MT: metatranscriptomes; TN: Total nitrogen; ENVI: Environment for Visualizing Images; BGI: Beijing Genomics Institute; NCBI: National Center for Biotechnology Information; SRA: Sequence Read Archive; QIIME: Quantitative Insights into Microbial Ecology; OTUs: Operational Taxonomic Units.

Declarations

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Availability of data and materials

Competing interests

The authors declare that they have no competing interests.

Funding

Authors' contributions

CMZ, JYZ, ZHL, NC and TC conceived the study and participated in its design. JYZ performed the sampling and geochemical measurements of samples. QYG extracted the RNA and DNA from samples. CMZ analyzed the sequencing data and performed the bioinformatics analysis. CMZ and TC performed manuscript preparation. All authors read and approved the final manuscript.

Acknowledgements

References

1. Paerl, H.W., et al., *Algal blooms: noteworthy nitrogen.* Science, 2014. **346**(6206): p. 175.
2. Paerl, H.W. and J. Huisman, *Blooms Like It Hot.* Science, 2008. **320**(5872): p. 57-58.
3. Qin, B., et al., *Environmental issues of Lake Taihu, China.* Hydrobiologia, 2007. **581**(1): p. 3-14.
4. Chen, Y., et al., *Long-term dynamics of phytoplankton assemblages : Microcystis-domination in Lake Taihu, a large shallow lake in China.* Journal of Plankton Research, 2003. **25**(4): p. 445-453(9).
5. Hai, X., et al., *Nitrogen and phosphorus inputs control phytoplankton growth in eutrophic Lake Taihu, China.* Limnology & Oceanography, 2010. **55**(1): p. 420-432.
6. Grant, M.A., et al., *Direct exchange of vitamin B12 is demonstrated by modelling the growth dynamics of algal-bacterial cocultures.* ISME J, 2014. **8**(7): p. 1418-27.
7. Shi, L., et al., *Specific association between bacteria and buoyant Microcystis colonies compared with other bulk bacterial communities in the eutrophic Lake Taihu, China.* Environ Microbiol Rep, 2012. **4**(6): p. 669-78.
8. Zhu, C., et al., *Seasonal succession and spatial distribution of bacterial community structure in a eutrophic freshwater Lake, Lake Taihu.* Sci Total Environ, 2019. **669**: p. 29-40.
9. Tang, X., et al., *Pyrosequencing analysis of free-living and attached bacterial communities in Meiliang Bay, Lake Taihu, a large eutrophic shallow lake in China.* Canadian Journal of Microbiology, 2015. **61**(1): p. 22-31.
10. Zhu, C.M., et al., *Alternate succession of aggregate-forming cyanobacterial genera correlated with their attached bacteria by co-pathways.* Sci Total Environ, 2019. **688**: p. 867-879.
11. Zhao, G., et al., *The importance of bacteria in promoting algal growth in eutrophic lakes with limited available phosphorus.* Ecological Engineering, 2012. **42**: p. 107-111.
12. Xie, M., et al., *Metagenomic Analysis Reveals Symbiotic Relationship among Bacteria in Microcystis-Dominated Community.* Frontiers in Microbiology, 2016. **7**(410).
13. Wang, W., et al., *Experimental evidence for the role of heterotrophic bacteria in the formation of Microcystis colonies.* Journal of Applied Phycology, 2015: p. 1-13.

14. Shia, L., et al., *Community Structure of Bacteria Associated with Microcystis Colonies from Cyanobacterial Blooms.* Journal of Freshwater Ecology, 2010. **25**(2): p. 193-203.

15. Zhu, C.-M., et al., *Alternate succession of aggregate-forming cyanobacterial genera correlated with their attached bacteria by co-pathways.* Science of The Total Environment, 2019. **688**.

16. Kiage, L.M. and N.D. Walker, *Using NDVI from MODIS to Monitor Duckweed Bloom in Lake Maracaibo, Venezuela.* Water Resources Management, 2009. **23**(6): p. 1125-1135.

17. Wei FS, Q.W., *Standard methods for water and wastewater monitoring and analysis.* China Environmental Science Press, 2002. **4th edn**.

18. Ding, J.Q. and J.Y. Zhang, *Application of YSI 6600 for Cyanobacteria Bloom Early Warning in Taihu Lake.* Administration & Technique of Environmental Monitoring, 2011.

19. Deben, C., et al., *Expression analysis on archival material revisited: isolation and quantification of RNA extracted from FFPE samples.* Diagnostic Molecular Pathology the American Journal of Surgical Pathology Part B, 2013. **22**(1): p. 59-64.

20. Caporaso, J.G., et al., *QIIME allows analysis of high-throughput community sequencing data.* Nat Methods, 2010. **7**(5): p. 335-6.

21. Desantis, T.Z., et al., *Greengenes, a chimera-checked 16S rRNA gene database and workbench compatible with ARB.* Applied & Environmental Microbiology, 2006. **72**(7): p.: 5069–5072.

22. O'Hara, R.B., et al., *vegan: community ecology package. Version 2.0-2.* Journal of Statistical Software, 2011. **48**(9): p. 1-21.

23. Simpson, G.L., et al., *Vegan: community ecology package.* Time International, 2015. **1997**(6): p. 15–17.

24. Li, D., et al., *MEGAHIT: an ultra-fast single-node solution for large and complex metagenomics assembly via succinct de Bruijn graph.* Bioinformatics, 2015. **31**(10): p. 1674-6.

25. Hyatt, D., et al., *Gene and translation initiation site prediction in metagenomic sequences.* Bioinformatics, 2012. **28**(17): p. 2223-2230.

26. Quinlan, A.R. and I.M. Hall, *BEDTools: a flexible suite of utilities for comparing genomic features.* Bioinformatics, 2010. **26**(6): p. 841-842.

27. Buchfink, B., C. Xie, and D.H. Huson, *Fast and sensitive protein alignment using DIAMOND.* Nat Meth, 2015. **12**(1): p. 59-60.

28. Huson, D.H., et al., *MEGAN analysis of metagenomic data.* Genome research, 2007. **17**(3): p. 377-386.

29. Kanehisa, M., et al., *The KEGG resource for deciphering the genome.* Nucleic Acids Res, 2004. **32**(Database issue): p. D277-80.

30. Overbeek, R., et al., *The subsystems approach to genome annotation and its use in the project to annotate 1000 genomes.* Nucleic Acids Res, 2005. **33**(17): p. 5691-702.

31. Alneberg, J., et al., *Binning metagenomic contigs by coverage and composition.* Nature methods, 2014. **11**(11): p. 1144-1146.

32. Li, H.J.B. and et al., *The Sequence Alignment / Map (SAM) Format.* 2009. **25**(1 Pt 2): p. 1653-4.

33. Parks, D.H., et al., *CheckM: assessing the quality of microbial genomes recovered from isolates, single cells, and metagenomes.* Genome research, 2015. **25**(7): p. 1043-1055.

34. Rahman, S.F., et al., *Machine Learning Leveraging Genomes from Metagenomes Identifies Influential Antibiotic Resistance Genes in the Infant Gut Microbiome.* 2018. **3**(1): p. e00123-17.

35. Na, S.-I., et al., *UBCG: Up-to-date bacterial core gene set and pipeline for phylogenomic tree reconstruction.* 2018.

36. Zhang, H., et al., *EvolView, an online tool for visualizing, annotating and managing phylogenetic trees.* 2012(W1): p. W1.

37. Grabherr, M.G., et al., *Grabherr, M.G.: Full-length transcriptome assembly from RNA-Seq data without a reference genome. Nat. Biotech. 29(7), 644-652.* Nature Biotechnology, 2011. **29**(7): p. 644-652.

38. Haas, B.J., et al., *De novo transcript sequence reconstruction from RNA-Seq: reference generation and analysis with Trinity.* Nature Protocols, 2013. **8**(8): p. 1494-1512.

39. Hultman, J., et al., *Multi-omics of permafrost, active layer and thermokarst bog soil microbiomes.* Nature, 2015. **521**(7551): p. 208-12.

40. Scholz, M., et al., *Strain-level microbial epidemiology and population genomics from shotgun metagenomics.* Nat Methods, 2016. **13**(5): p. 435-8.

41. Parks, D.H., et al., *STAMP: statistical analysis of taxonomic and functional profiles.* Bioinformatics, 2014. **30**(21): p. 3123.

42. Lauro, F.M., et al., *The genomic basis of trophic strategy in marine bacteria.* Proceedings of the National Academy of Sciences of the United States of America, 2009. **106**(37): p. 15527-15533.

43. Tilman, D., *Resource competition and community structure.* Monographs in Population Biology, 1982. **17**(17): p. 1.

44. Bowers, R.M., et al., *Minimum information about a single amplified genome (MISAG) and a metagenome-assembled genome (MIMAG) of bacteria and archaea.* Nature biotechnology, 2017. **35**(8): p. 725-731.

45. Meierkolthoff, J.P, et al., *Genome sequence-based species delimitation with confidence intervals and improved distance functions.* Bmc Bioinformatics, 2013. **14**(1): p. 60.

46. Zhang, D., et al., *Spatial and temporal variations of microcystins in hepatopancreas of a freshwater snail from Lake Taihu.* Ecotoxicol Environ Saf, 2009. **72**(2): p. 466-72.

47. Niu, Y., et al., *Phytoplankton community succession shaping bacterioplankton community composition in Lake Taihu, China.* Water Res, 2011. **45**(14): p. 4169-82.

48. Ye, W., et al., *Temporal variability of cyanobacterial populations in the water and sediment samples of Lake Taihu as determined by DGGE and real-time PCR.* Harmful Algae, 2011. **10**(5): p. 472-479.

49. Tan, X., et al., *Seasonal variation of Microcystis in Lake Taihu and its relationships with environmental factors.* Journal of Environmental Sciences, 2009. **21**(7): p. 892-899.

50. Cai, H.Y., et al., *Analysis of the attached microbial community on mucilaginous cyanobacterial aggregates in the eutrophic Lake Taihu reveals the importance of Planctomycetes.* Microbial Ecology, 2013. **66**(1): p. págs. 73-83.

51. Zhang, J., et al., *Microbial profiles of a drinking water resource based on different 16S rRNA V regions during a heavy cyanobacterial bloom in Lake Taihu, China.* Environ Sci Pollut Res Int, 2017. **24**(14): p. 12796-12808.

52. Mcbride, M.J., et al., *The Family Cytophagaceae.* 2014: Springer Berlin Heidelberg. 577-593.

53. Romanovicz, D.K. and R.M. Brown, *Cellulose in Cyanobacteria. Origin of Vascular Plant Cellulose Synthase?* Plant Physiology, 2001. **127**(2): p. 529-542.

54. Park, C.B., et al., *An Eco-Safety Assessment of Glyoxal-Containing Cellulose Ether on Freeze-Dried Microbial Strain, Cyanobacteria,Daphnia, and Zebrafish.* International Journal of Environmental Research & Public Health, 2017. **14**(3): p. 323.

55. Fernández, C., V. Estrada, and E.R. Parodi, *Factors Triggering Cyanobacteria Dominance and Succession During Blooms in a Hypereutrophic Drinking Water Supply Reservoir.* Water Air & Soil Pollution, 2015. **226**(3): p. 1-13.

56. Wu, W., et al., *Temperature may be the dominating factor on the alternant succession of Aphanizomenon flos-aquae and Microcystis aeruginosa in Dianchi Lake.* Fresenius Environmental Bulletin, 2010. **19**(5): p. 846-853.

57. Sauer, K., et al., *Pseudomonas aeruginosa displays multiple phenotypes during development as a biofilm.* Journal of Bacteriology, 2002. **184**(4): p. 1140-1154.

58. Koehorst, J.J., et al., *Comparison of 432 Pseudomonas strains through integration of genomic, functional, metabolic and expression data.* Scientific Reports, 2016. **6**: p. 38699.

59. Hassett, D.J., et al., *Anaerobic metabolism and quorum sensing by Pseudomonas aeruginosa biofilms in chronically infected cystic fibrosis airways: rethinking antibiotic treatment strategies and drug targets.* Advanced Drug Delivery Reviews, 2002. **54**(11): p. 1425-1443.

60. Fera, M.T., et al., *Detection of Arcobacter spp. in the coastal environment of the Mediterranean Sea.* Applied & Environmental Microbiology, 2004. **70**(3): p. 1271-6.

61. Bender, K.S., et al., *Sequencing and Transcriptional Analysis of the Chlorite Dismutase Gene of Dechloromonas agitata and Its Use as a Metabolic Probe.* Applied & Environmental Microbiology, 2002. **68**(10): p. 4820-6.

62. LC, M., et al., *An epidemic, toxin gene-variant strain of Clostridium difficile.* N Engl J Med, 2005. **353**(23): p. 2433-2441.

63. Baban, S.T., et al., *The Role of Flagella in Clostridium difficile Pathogenesis: Comparison between a Non-Epidemic and an Epidemic Strain.* Plos One, 2013. **8**(9): p. e73026.

64. Willem, A., *Comamonadaceae, a new family encompassing the acidovorans rRNA complex, including Variovorax paradoxus gen. nov., com. nov., for Alcaligenes paradoxus (Davis 1969).* Int.j.syst.bacteriol, 1991. **41**(3): p. 445-450.

65. Adam, B., et al., *N2-fixation, ammonium release and N-transfer to the microbial and classical food web within a plankton community.* Isme Journal, 2015. **10**(2).

66. Bar-Yosef, Y., et al., *Enslavement in the Water Body by Toxic Aphanizomenon ovalisporum, Inducing Alkaline Phosphatase in Phytoplanktons.* Current Biology, 2010. **20**(17): p. 1557-1561.

67. Takakazu, K., et al., *Complete Genomic Structure of the Bloom-forming Toxic Cyanobacterium Microcystis aeruginosa NIES-843.* DNA Research, 2008. **14**(6): p. 247-256.

68. Karl, D.M., *Nutrient dynamics in the deep blue sea.* Trends Microbiol, 2002. **10**(9): p. 410-8.

69. Haft, D.H., et al., *A guild of 45 CRISPR-associated (Cas) protein families and multiple CRISPR/Cas subtypes exist in prokaryotic genomes.* PLoS computational biology, 2005. **1**(6): p. e60-e60.

70. Brouns, S.J.J., et al., *Small CRISPR RNAs Guide Antiviral Defense in Prokaryotes.* ence, 2008. **321**(5891): p. p.960-964.

71. Scholz, I., et al., *CRISPR-Cas systems in the cyanobacterium Synechocystis sp. PCC6803 exhibit distinct processing pathways involving at least two Cas6 and a Cmr2 protein.* PLoS One, 2013. **8**(2): p. e56470.

72. Bbosa, N.B. and W.S. Oyoo, *Seasonal variations of phytoplankton species in Lake Victoria and the influence of iron and zinc ions on the dominant species identified during 2006–2007 studies.* Lakes & Reservoirs Research & Management, 2013. **18**(3): p. 259–273.

73. Wagner-Döbler, I., et al., *The complete genome sequence of the algal symbiont Dinoroseobacter shibae: a hitchhiker's guide to life in the sea.* Isme j, 2010. **4**(1): p. 61-77.

Tables

Table 1. Sample sequencing data and sample grouping.

Stage	Sample ID	Bloom area (km^2)	16S (Mb)	MG (Gb)	MT (Gb)
Early	150402	2.42	13.72	9.77	4.85
	150409	3.26	13.73	9.77	7.06
	150514	7.69	14.05	9.63	10.1
Mid	150608	102.55	28.78	6.58	6.89
	150615	135.8	28.94	6.48	9.59
	150701	232.09	29.07	6.59	13.1
	150804	23.85	29.07	6.66	9.78
	150806	59.36	28.60	6.63	10.2
	150902	19.86	28.54	6.63	10.5
	151028	12.18	27.45	6.75	10.7
	151123	481.10	27.70	6.66	12.6
	151211	66.71	28.22	6.78	9.99
	151222	157.67	27.48	6.53	10.2
Late	160104	1.13	27.88	6.71	8.2
	160119	1.65	27.60	6.60	11.1
	160127	2.31	28.02	6.78	8.49

Table 2. Top 10 genera with highest activity (MG/MT ratio) for three stages.

Stages	Genus	Activity
Early	Pseudomonas	3.75356
	Dolichospermum	3.169236
	Pseudanabaena	2.905012
	Leptospira	2.735959
	unclassified Proteobacteria	2.081191
	Anabaena	1.924466
	Aphanizomenon	1.877147
	Microcystis	1.713039
	unclassified Cytophagaceae	1.627376
	Hassallia	1.496396
Mid	Arcobacter	12.5225
	Dechloromonas	5.733142
	Clostridium	4.505952
	Pseudomonas	3.746735
	Desulfovibrio	1.71246
	Flavobacterium	1.679435
	Rickettsia	1.604167
	Thauera	1.431159
	Pseudanabaena	1.38034
	Herbaspirillum	1.130303
Late	Dolichospermum	4.39707
	Aphanizomenon	4.039871
	Pseudomonas	4.035642
	Anabaena	2.563843
	Rubrivivax	1.831885
	unclassified Burkholderiales (miscellaneous)	1.558568
	Arcobacter	1.551598
	Azohydromonas	1.535461
	Caldimonas	1.532407
	Herbaspirillum	1.492691

Table 3. Basic information of 78 non-redundant MAGs.

MAGs ID	Com.%	Con.%	Len.Mb	Phylum	MAGs ID	Com.%	Con.%	Len.Mb	Phylum
150402.36	95.9	1.33	3.43	Proteobacteria	150804.24	96.0	2.19	3.54	Proteobacteria
150402.45	89.7	4.75	3.43	Proteobacteria	150804.57	100	4.62	3.67	Unkown
150402.69	97.3	1.81	3.62	Bacteroidetes	150804.71	90.1	0.65	1.58	Proteobacteria
150409.101	98.2	1.85	3.99	Proteobacteria	150804.78	98.6	0.82	3.59	Proteobacteria
150409.124	97.4	3.57	1.72	Proteobacteria	150804.9	95.5	0	3.83	Proteobacteria
150409.14	97.0	2.53	4.45	Bacteroidetes	150804.93	95.1	0.62	4.81	Proteobacteria
150409.20	98.1	1.18	3.35	Proteobacteria	150804.95	98.3	0.83	3.68	Proteobacteria
150409.60	95.1	4.72	3.63	Proteobacteria	150806.10	98.3	1.81	3.26	Proteobacteria
150409.80	99.1	0	2.64	Bacteroidetes	150806.104	98.4	0.97	3.83	Proteobacteria
150409.86	90.5	3.44	5.13	Unkown	150806.13	95.8	1.4	3.40	Bacteroidetes
150409.99	96.5	0	3.70	Spirochaetes	150806.14	96.5	0.92	5.25	Acidobacteria
150514.103	96.0	0.57	4.08	Unkown	150806.23	96.0	3.84	3.06	Proteobacteria
150514.116	98.5	2.42	2.64	Proteobacteria	150806.32	73.6	2.04	3.12	Proteobacteria
150514.122	90.7	0.51	2.33	Unkown	150806.49	91.2	2.34	8.71	Proteobacteria
150514.128	93.6	0.16	4.92	Proteobacteria	150806.50	78.9	0.65	2.42	Unkown
150514.139	97.8	4.4	3.14	Proteobacteria	150806.69	94.4	2.52	3.26	Proteobacteria
150514.36	70.2	3.58	3.43	Proteobacteria	150806.71	98.9	3.85	4.65	Gemmatimonadetes
150514.92	95.4	4.59	3.76	Proteobacteria	150806.73	94.9	1.31	2.94	Unkown
150608.53	97.1	1.6	2.86	Proteobacteria	150806.95	87.7	2.63	3.14	Proteobacteria
150608.54	72.1	2.48	2.73	Proteobacteria	150902.3	97.7	0	3.84	Unkown
150608.55	94.9	2.44	3.86	Bacteroidetes	151028.46	91.2	0	2.25	Unkown
150608.59	99.6	0.45	4.68	Cyanobacteria	151028.71	81.7	3.82	2.97	Proteobacteria
150608.83	91.9	3.97	3.67	Proteobacteria	151028.73	99.1	0.54	3.39	Bacteroidetes
150615.15	98.4	3.48	4.02	Proteobacteria	151028.77	98.7	1.65	4.47	Cyanobacteria
150615.22	92.7	2.83	2.28	Proteobacteria	151123.13	95.8	1.79	1.83	Bacteroidetes
150615.30	99.7	2.19	3.35	Bacteroidetes	151123.29	97.1	3.77	2.36	Bacteroidetes
150615.38	71.1	0.29	2.32	Proteobacteria	151123.5	88.6	0	2.14	Proteobacteria
150615.43	83.7	0.1	1.52	Proteobacteria	151123.70	99.6	0.16	2.72	Proteobacteria
150615.60	71.4	4.84	3.37	Proteobacteria	151123.81	89.0	4.56	2.51	Bacteroidetes
150615.66	97.8	2.86	2.93	Bacteroidetes	151211.17	96.0	3.54	3.54	Proteobacteria
150701.27	95.1	4.46	3.70	Bacteroidetes	151211.78	93.8	1.65	3.57	Unkown
150701.64	94.3	1.7	3.76	Unkown	151222.16	91.7	4.47	3.68	Proteobacteria
150701.80	92.4	4.22	2.35	Proteobacteria	151222.71	97.5	1.58	3.91	Proteobacteria
150804.103	95.2	2.3	4.75	Gemmatimonadetes	160104.35	98.2	1.35	3.30	Bacteroidetes
150804.108	98.3	0.65	3.42	Proteobacteria	160104.55	84.4	3.45	3.99	Proteobacteria
150804.112	99.0	0.65	4.73	Proteobacteria	160104.80	73.4	0.86	3.13	Bacteroidetes
150804.114	94.5	1.35	3.07	Proteobacteria	160119.20	97.9	1.22	3.04	Proteobacteria

| 150804.12 | 97.1 | 1.87 | 4.11 | Unkown | 160127.41 | 95.3 | 2.08 | 3.79 | Bacteroidetes |
| 150804.17 | 96.6 | 0.07 | 3.40 | Unkown | 160127.48 | 80.9 | 2.5 | 2.68 | Spirochaetes |

Com.: Completeness; Con.: Contamination; Len.: Length.
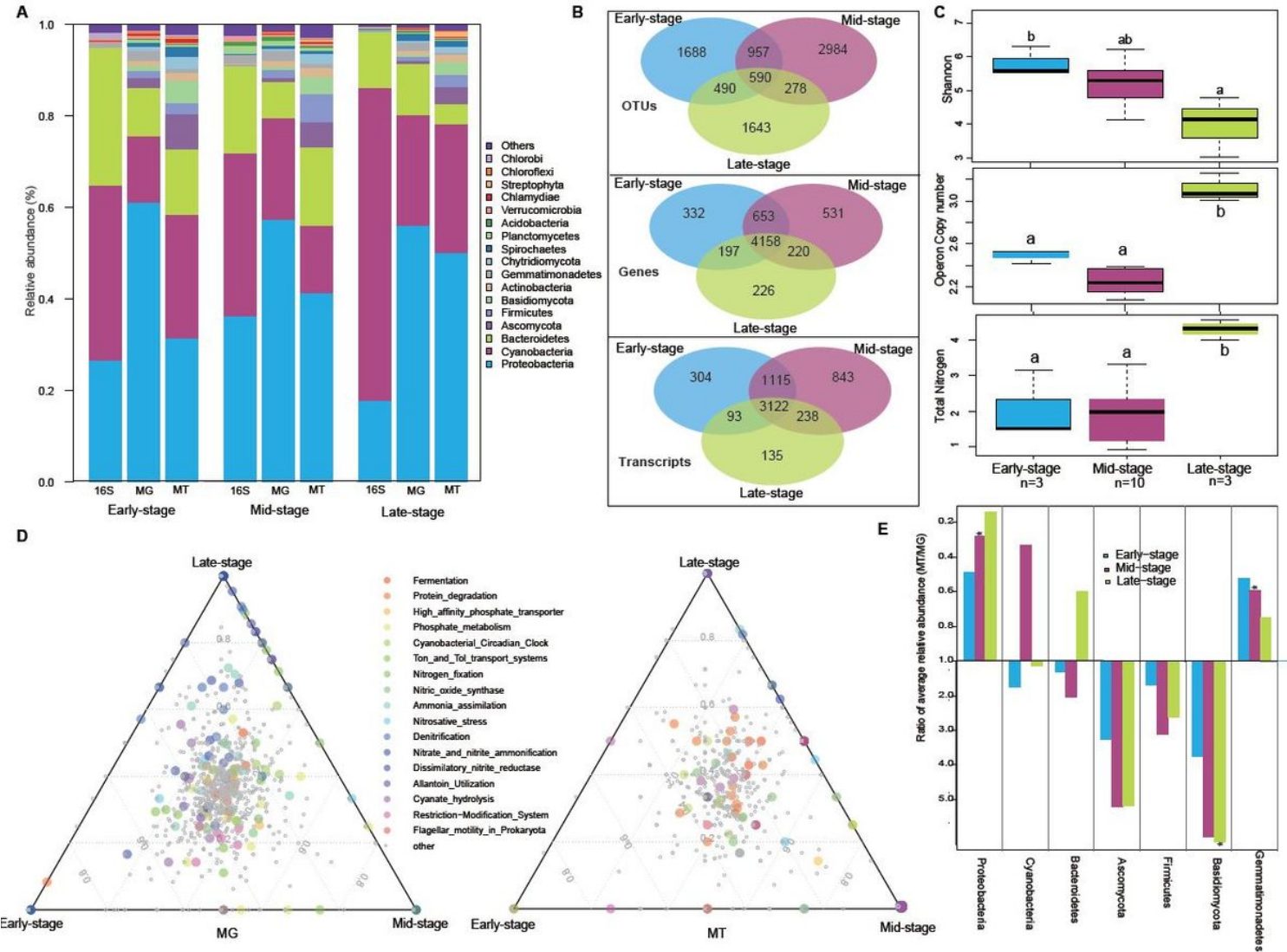
Figures

Figure 1

Microbial taxonomic and functional composition in CAs of three different stages (A) Relative abundance of phyla on the basis of targeted 16S rRNA gene sequencing (16S), metagenomes (MG) and metatranscriptomes (MT). (B) Venn diagrams illustrate the unique and shared OTUs, genes and transcripts among three stages. (C) Comparison of alpha diversity, average rRNA copy number, total nitrogen among three stages. (D) Comparison of genes and transcripts among three stages. (E) Ratio of average relative abundance of seven abundant phyla in the metatranscriptomes relative to the metagenomes (MT/MG). Bars marked with an asterisk are significantly different (two-tailed t-test, P value<0.05) from MT/MG=1.0.
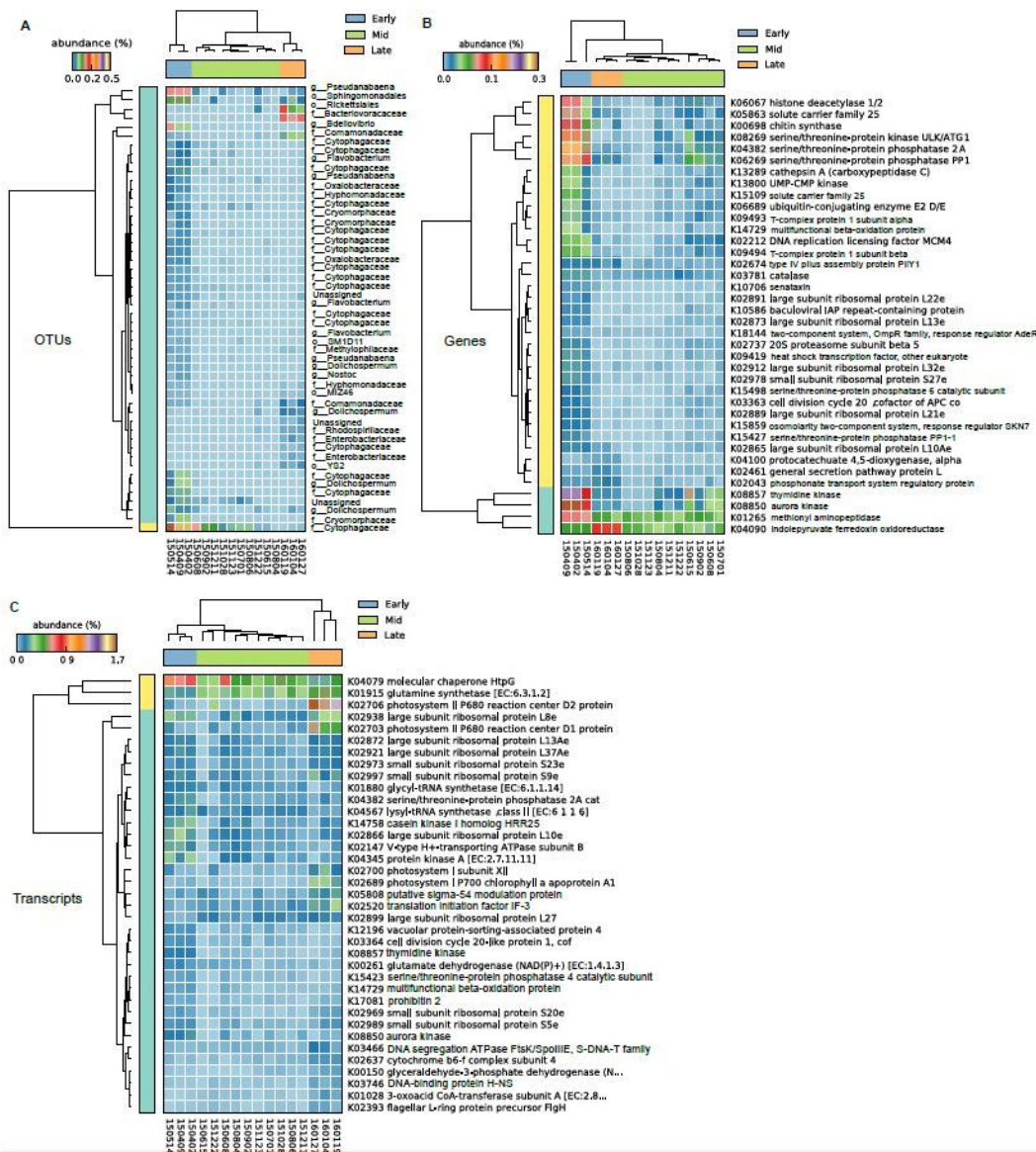
Figure 2

OTUs, genes and transcripts with significantly different abundance among three stages via ANOVA test (FDR<0.05). (A) 55 OTUs. Each line represents an OTU annotated with GreenGene database. Each column represents a sample. (B) 38 genes. Each line represents a gene annotated with KEGG database. (C) 37 transcripts. Each line represents a transcript annotated with KEGG database.
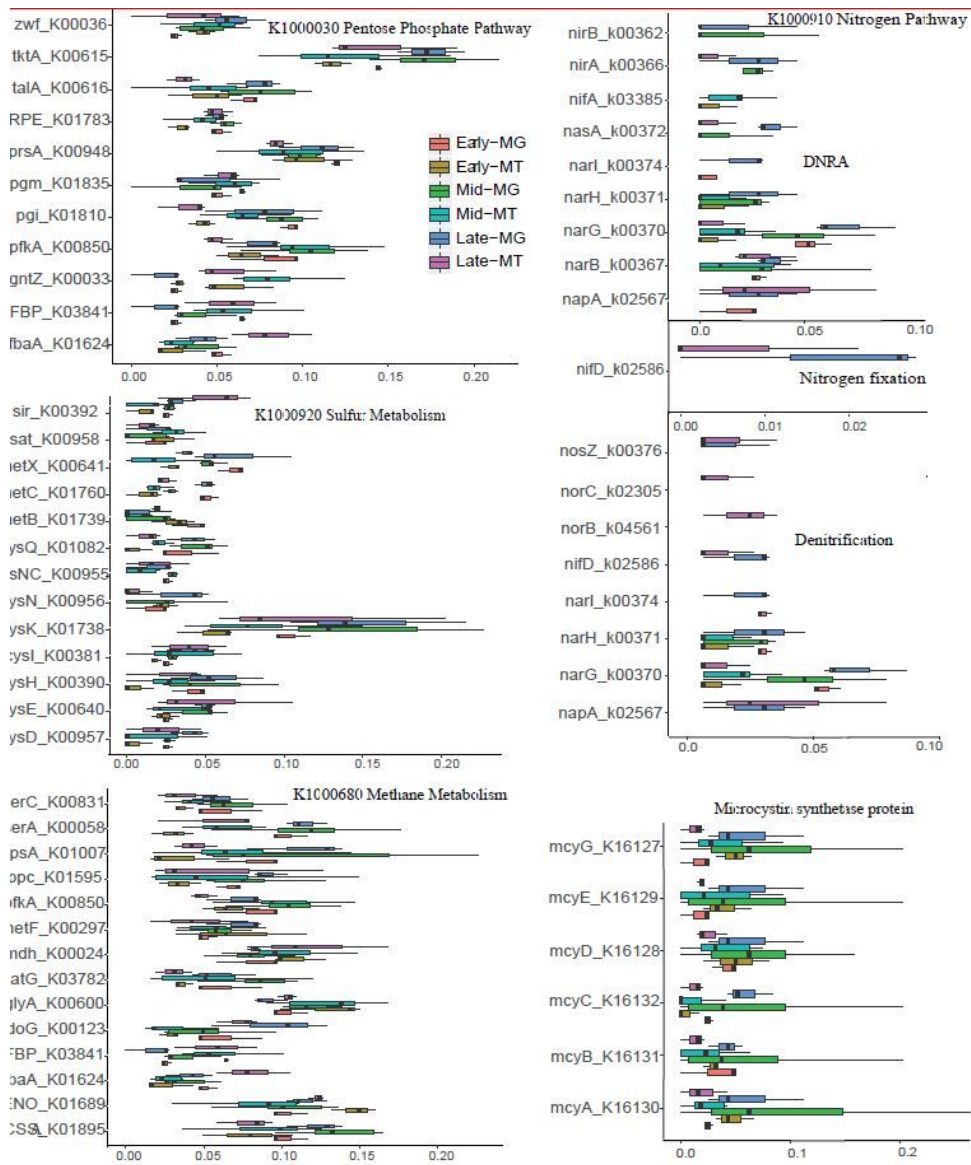
Figure 3

Abundance of genes involved in pentose phosphate pathway, sulfur metabolism, methane metabolism, nitrogen cycle and microcystin synthetase. The relative abundance of each gene is shown; error bars, s.d. The genes with protein match in the metaproteome data are also showed.
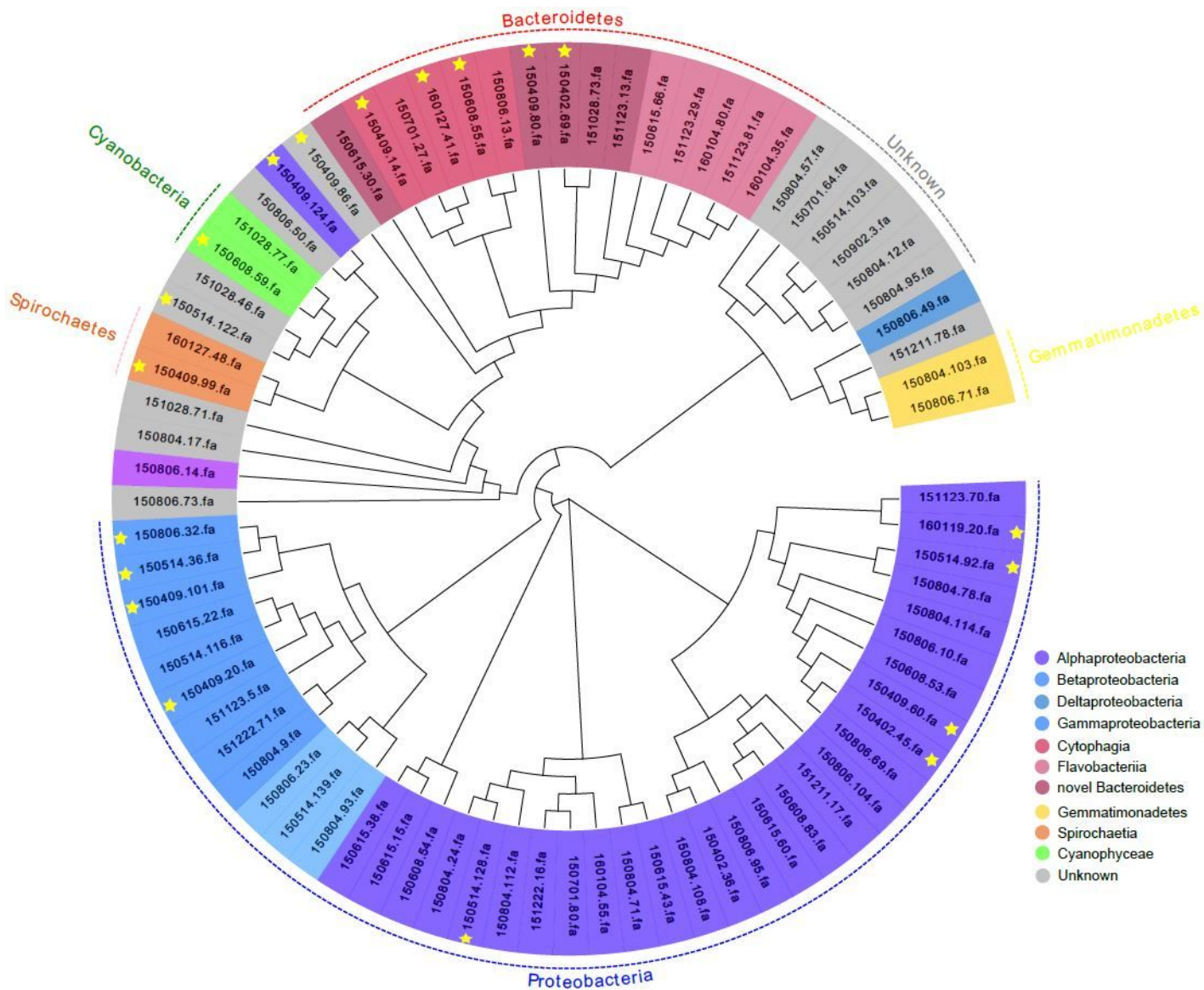
Figure 4

Phylogenetic tree of 78 non-redundant MAGs. The tree was produced from bacterial core gene set using UBCG, and subsequently drawn using Evolview. Labels show MAGs' names with the sample name and the colors show the class which the MAGs belong to. The 19 MAGs labeled with yellow stars showed significantly different abundance among three stages.
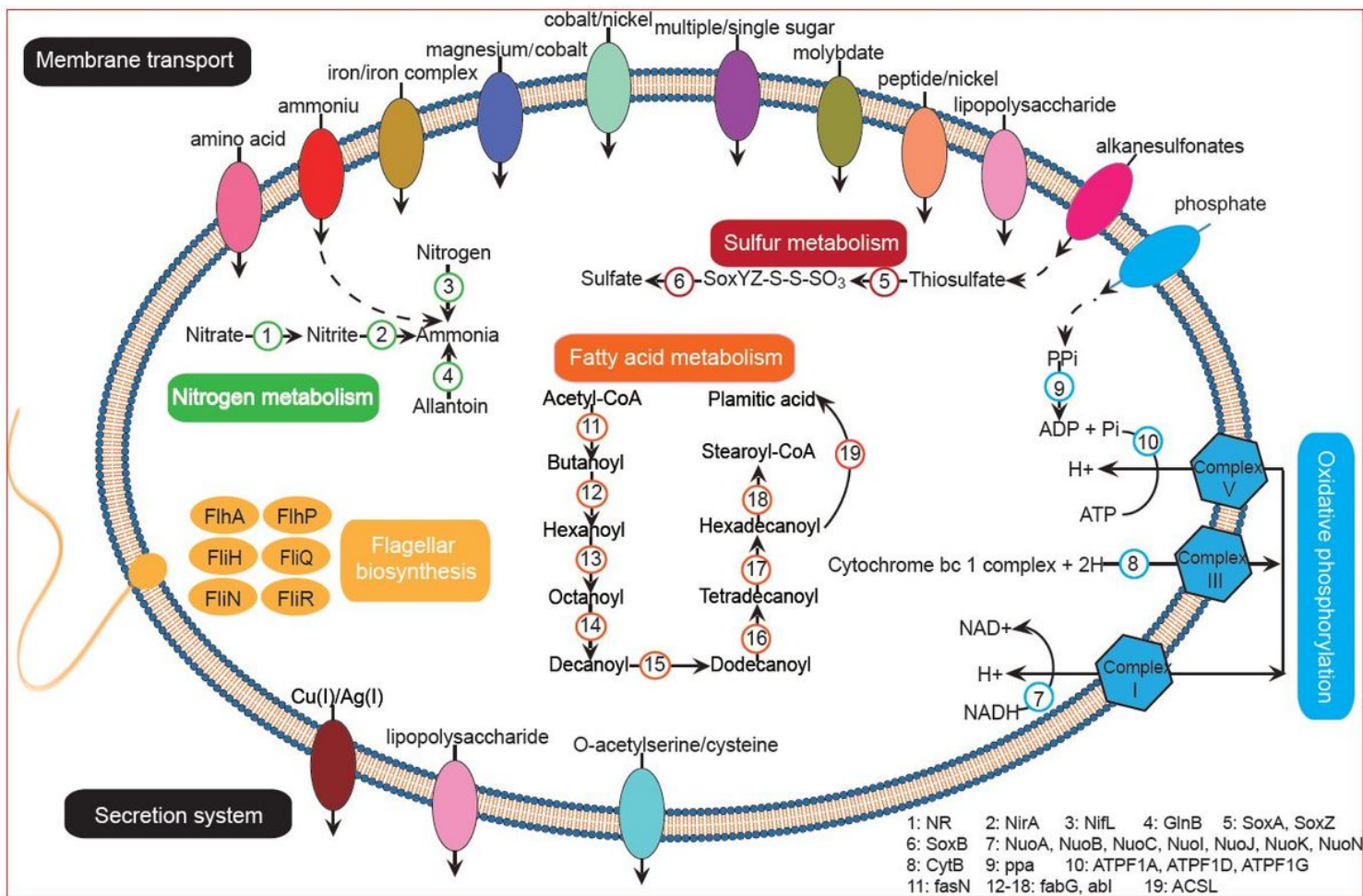
Figure 5

Visualization of metabolic pathways of the most abundant MAG. The most abundant MAG (151222.71.fa) has rich membranetransporters (represented with ellipses on the membrane) and metabolic pathways including flagellar biosynthesis pathway (yellow), nitrogen metabolism transferring nitrate to ammonia (green), sulfur metabolism transferring thiosulfate to sulfate (red), the fatty acid metabolism transferring acetyl-coa to plamitic acid (orange) and the oxidative phosphorylation pathway releasing energy supplies ADP and inorganic phosphoric acid to synthesize ATP through respiratory chain (blue).
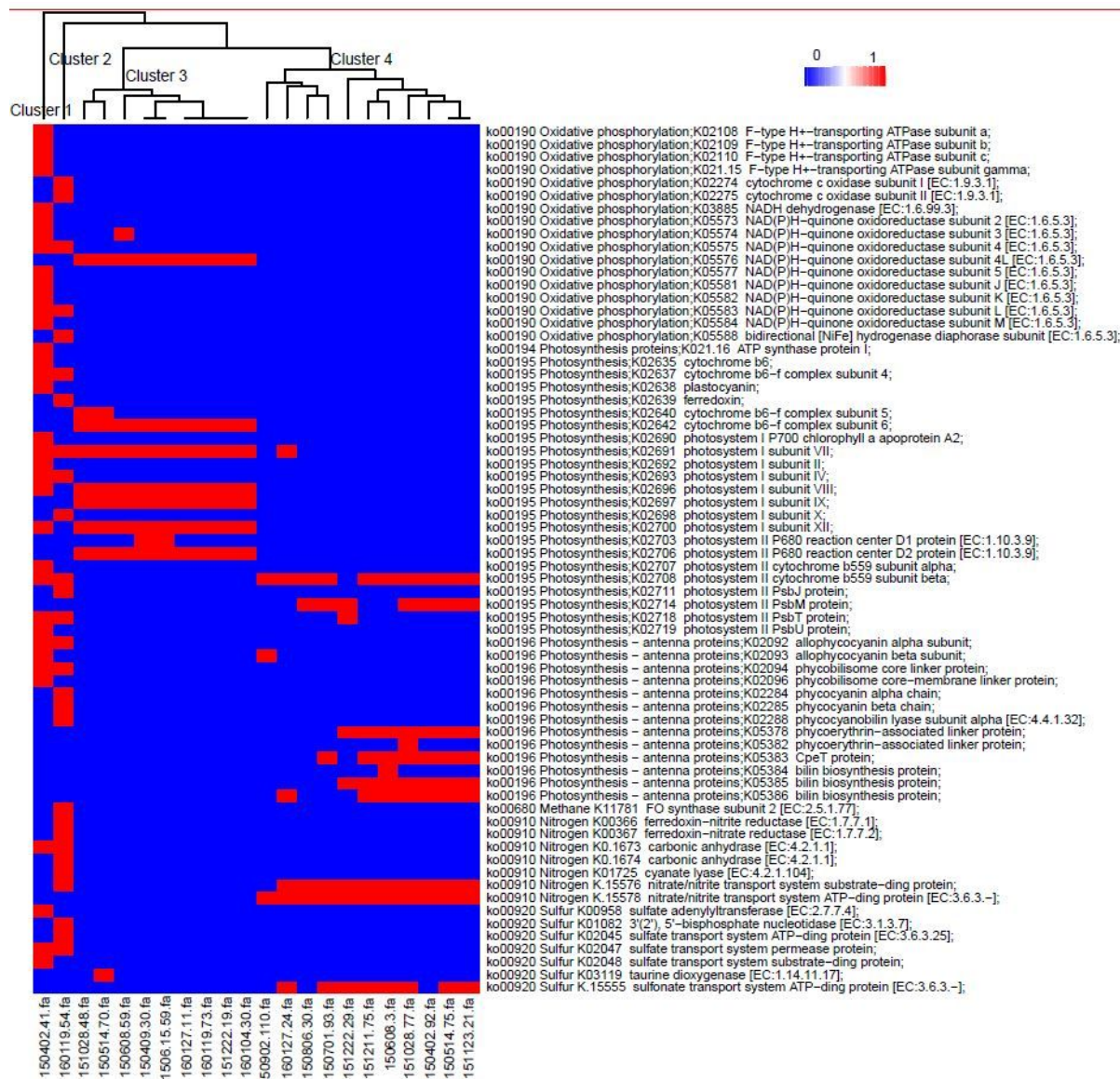
Figure 6

Heatmap of genes on energy pathways of 22 cyanobacterial MAGs. Genes on energy pathways of nitrogen metabolism, sulfur metabolism and photosynthesis. Each column represents a cyanobacterial MAG and each row represents a gene annotated with KEGG.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- Additionalfile7.txt
- Additionalfile6.pdf
- Additionalfile5.pdf
- Additionalfile4.pdf
- Additionalfile3.pdf
- Additionalfile2.pdf
- Additionalfile1.pdf