



Metagenomic Analysis Reveals Symbiotic Relationship among Bacteria in *Microcystis*-Dominated Community

Meili Xie^{1,2†}, Minglei Ren^{1,2†}, Chen Yang^{1,2}, Haisi Yi^{1,2}, Zhe Li^{3*}, Tao Li^{1*} and Jindong Zhao^{1,4}

¹ Key Laboratory of Algal Biology, Institute of Hydrobiology, Chinese Academy of Sciences, Wuhan, China, ² University of Chinese Academy of Sciences, Beijing, China, ³ State Key Laboratory of Systematic and Evolutionary Botany, Institute of Botany, Chinese Academy of Sciences, Beijing, China, ⁴ College of Life Science, Peking University, Beijing, China

OPEN ACCESS

Edited by:

Petra M. Visser,
University of Amsterdam, Netherlands

Reviewed by:

Hans-Peter Grossart,
IGB-Leibniz-Institute of Freshwater
Ecology and Inland Fisheries,
Germany
Haiwei Luo,
The Chinese University of Hong Kong,
China

*Correspondence:

Zhe Li
lizhe@ibcas.ac.cn;
Tao Li
lita@ihb.ac.cn

[†]These authors have contributed
equally to this work.

Specialty section:

This article was submitted to
Aquatic Microbiology,
a section of the journal
Frontiers in Microbiology

Received: 02 November 2015

Accepted: 13 January 2016

Published: 02 February 2016

Citation:

Xie M, Ren M, Yang C, Yi H, Li Z, Li T
and Zhao J (2016) Metagenomic
Analysis Reveals Symbiotic
Relationship among Bacteria in
Microcystis-Dominated Community.
Front. Microbiol. 7:56.
doi: 10.3389/fmicb.2016.00056

Microcystis bloom, a cyanobacterial mass occurrence often found in eutrophicated water bodies, is one of the most serious threats to freshwater ecosystems worldwide. In nature, *Microcystis* forms aggregates or colonies that contain heterotrophic bacteria. The *Microcystis*-bacteria colonies were persistent even when they were maintained in lab culture for a long period. The relationship between *Microcystis* and the associated bacteria was investigated by a metagenomic approach in this study. We developed a visualization-guided method of binning for genome assembly after total colony DNA sequencing. We found that the method was effective in grouping sequences and it did not require reference genome sequence. Individual genomes of the colony bacteria were obtained and they provided valuable insights into microbial community structures. Analysis of metabolic pathways based on these genomes revealed that while all heterotrophic bacteria were dependent upon *Microcystis* for carbon and energy, Vitamin B12 biosynthesis, which is required for growth by *Microcystis*, was accomplished in a cooperative fashion among the bacteria. Our analysis also suggests that individual bacteria in the colony community contributed a complete pathway for degradation of benzoate, which is inhibitory to the cyanobacterial growth, and its ecological implication for *Microcystis* bloom is discussed.

Keywords: metagenome, binning, *Microcystis*, bloom, symbiosis

INTRODUCTION

Cyanobacterial blooms are mass occurrence of cyanobacterial species in water bodies and are serious threats to freshwater ecosystems worldwide (Bláha et al., 2009). Due to the increased input of nutrients into lakes and rivers, eutrophication has become a global issue and it is recognized as one of the most important factors in the formation of cyanobacterial blooms in freshwater ecosystems. Among blooms of the several cyanobacterial genera, *Microcystis* blooms might be the most widely spread globally. The problems caused by the *Microcystis* blooms include the release of toxic secondary metabolites such as microcystins into the water environments that cause intoxication of humans and animals (Sivonen and Jones, 1999; Falconer and Humpage, 2006) and the formation of hypoxic zones in water bodies that lead to death of fish and other organisms.

Accumulation of cyanobacteria and release of odorous materials also cause technical problems in water treatment plants and dramatically decrease the drinking water quality.

The formation of cyanobacterial blooms has been studied extensively and much progress has been made in understanding the relationships between the blooms and environmental factors (Conley et al., 2009; Davis et al., 2009; Ma et al., 2014). Recently, people have started to pay more attention to the importance of symbiotic relationship between *Microcystis* and the associated heterotrophic bacteria (Eiler and Bertilsson, 2004; Sigeo, 2005; Berg et al., 2008; Dziallas and Grossart, 2011, 2012; Shen et al., 2011). It is desirable to obtain the genomes of these heterotrophic bacteria and understand their ecological roles in *Microcystis* colonies. However, because only a small portion of bacteria in nature can be cultured in the lab and methods are quite limited in isolating individual genomic DNA or RNA samples, the complex structures of microbial interactions, including *Microcystis*-bacteria associations, are not well understood. The successful application of metagenomics in studies of microbial communities, such as ocean water, natural acidophilic biofilm, permafrost, acetate-amended aquifers, activated and sludge bioreactors, has led to new insights and mechanisms for studying bacterial communities in natural environments (Tyson et al., 2004; Venter et al., 2004; Mackelprang et al., 2011; Wrighton et al., 2012; Albertsen et al., 2013). The metagenomics approach has also been applied to cyanobacterial blooms (Pope and Patel, 2008; Eiler et al., 2011; Li et al., 2011; Steffen et al., 2012; Mou et al., 2013). These works have provided very useful information on the microbial communities in *Microcystis* blooms. However, due to the lack of whole genome information from the individual members of the community, details on the metabolic pathways and the relationships among members of the community are not well understood.

In metagenomic analysis, grouping sequences from a particular genome from microbial community sequencing data is an important step referred to as binning (Kunin et al., 2008; Mande et al., 2012). The binning process can greatly reduce the complexity of metagenomics data by grouping similar sequences together followed by assembly and annotation to the individual genome bins. Therefore, it enables researchers to decipher the functional roles of community members and their interactions. Currently, there are several programs designed to classify sequencing reads (Huson et al., 2007; Wu and Ye, 2011; Wang et al., 2012) from either metagenomic samples or assembled scaffolds (Dick et al., 2009; Albertsen et al., 2013). Recently, model-based clustering techniques, in particular the multi-component Gaussian mixture model, have been employed to bin metagenomic fragments (Alneberg et al., 2014; Laczny et al., 2014; Wu et al., 2014).

Here, we present a novel visualization-guided and model-based binning method. It was applied to the metagenomic dataset generated from a *Microcystis*-dominated bloom in Lake Taihu. The high-resolution binning results show that there were at least seven bacteria species in stable cohabitation with *Microcystis* species. The functional annotation and pathway analysis provided an insight into the relationships among *Microcystis* and the associated bacteria. It suggests that metabolic

pathway complementation plays an important role in stable colony formation and the long-term domination of *Microcystis* blooms. A possible impact of *Microcystis*-bacteria association on health effects for both humans and animals is discussed.

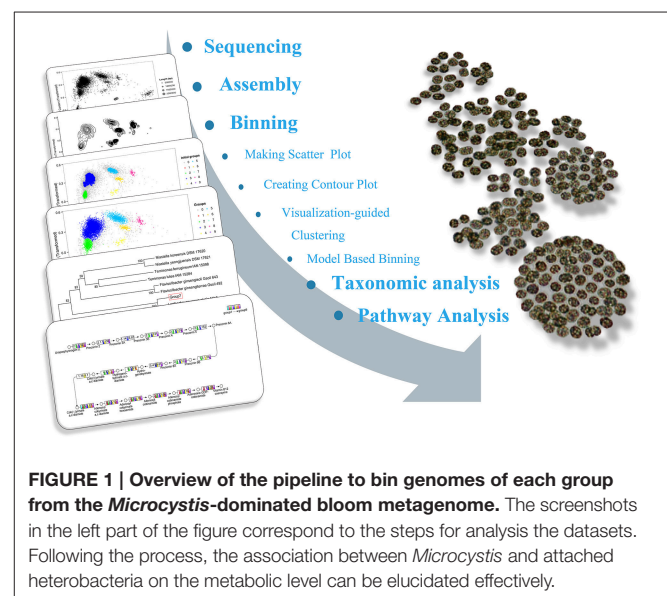
MATERIALS AND METHODS

Strains

Microcystis wesenbergii T100 was isolated from Lake Taihu in 2006, a freshwater shallow lake in East China in which serious cyanobacterial blooms occur annually. Non-microcystin-producing *M. wesenbergii* T100 forms large mucilaginous colonies that have embedded several species of bacteria. The cultures were maintained in BG-11 medium in an incubation room at 25°C under 25 μM photons $\text{m}^{-2} \text{s}^{-1}$ on a 12:12 light-dark cycle. A flow diagram of the binning and downstream analysis process was shown in Figure 1, and a step-by-step guide is available at <http://mingleir.github.io/meta-binning/>.

Metagenomics DNA Sequencing

In 2009, the cells were harvested in the exponential phase by filtering water samples through a 0.45- μm pore filter and disrupted with Mini-Beadbeater (BioSpec Products, Elgin, IL, USA) at maximum speed. Total DNA was extracted according to Li et al. (2001) with modifications. Briefly, the sample was suspended in TE buffer (pH 8.0) with 0.5% SDS and 1 mg/mL lysozyme incubated at 37°C for 1 h. Before digesting at 55°C for 2 h with 0.4 mg/mL proteinase K, 1% β -mercaptoethanol was added to prevent pigment oxidation. The mixture was purified with phenol-chloroform-isoamyl alcohol (25:24:1) and washed with 70% ethanol twice. The final supernatant was precipitated with isopropanol and dissolved in sterile water. A 300-bp paired-end library and a 3-Kb mate-paired library were constructed according to the manufacturer's instructions (www.Illumina.com). The libraries were sequenced on an



Illumina Genome Analyzer IIx following the manufacturer's instructions with 2×81 bp read length.

Data Preprocessing and Assembly

Quality trimming was performed in Trimmomatic v0.32 (Bolger et al., 2014) with reads shorter than 20 bp being discarded. After filtering, we used SPAdes v3.0.0 (Bankevich et al., 2012) to assemble the metagenomics data of cyanobacterial blooms in Lake Taihu with different parameters. Optional parameters were selected by considering the scaffolds *N50* statistic, the number of scaffolds, and the maximum scaffold length.

Data Visualization by Scatter Points

In the present study, we considered both sequence composition and relative abundance in the community in the assembled scaffold cluster analysis. GC content is the most fundamental compositional feature of a genome and has long been used in the identification of unknown DNA fragments (Karlin, 1998). The most intuitive measure of abundance is coverage. Sequences belonging to the same source population should have similar coverage (Strous et al., 2012). Recently, inspired by the rapid progress of RNA-sequencing, many statistical models have been proposed to accurately quantify transcript relative abundances (Pachter, 2011). Furthermore, the problem of estimating genome relative abundances in a community is closely related to relative transcript abundance estimation (Pachter, 2011). Fragments per kilo base per million mapped reads (FPKM) (Trapnell et al., 2010), defined as the read coverage normalized by the total number of mapped reads and sequence lengths, are widely accepted as a measure of quantifying relative transcript abundance and we used FPKM as a proxy for relative genome abundance in a community. We used Bowtie v. 1.0.1 (Langmead et al., 2009) to count the reads mapped on the scaffolds and eXpress v. 1.5.1 (Roberts and Pachter, 2013) to estimate the FPKM of each scaffold.

To show the structure in the set of sequence features, two-dimensional kernel density estimation was applied to the joint density function of GC content and FPKM. Each scaffold was weighted by its length because we investigated the distribution of both features at the nucleotide level. Because bandwidth is critical in the practical implementation of kernel density estimation (KDE) (Rosenblatt, 1956; Parzen, 1962), a two-stage direct plug-in bandwidth selector (Sheather and Jones, 1991; Wand and Jones, 1995) was used to produce a bandwidth estimate for the data set. We observed that sequence coverage contained outliers with extremely large values, which can have considerable effects on bandwidth selection. Therefore, we discarded scaffolds with coverage >0.99 quantile. Furthermore, logit transformation was applied to GC content, which was represented as proportion data, so that it follows a normal distribution. Binned kernel density estimation was applied to the transformed data (Wand, 1994; Wand and Jones, 1995). The standard bivariate normal density was used as the kernel. The estimated density function was visualized by contour plots. Guided by the graphs, the correct density function value was selected to initially bin the scaffolds.

Model-Based Clustering

From the initial binning graph described above, two facts could be observed: (1) each group could be modeled by a bivariate normal distribution; (2) a large number of data points (scaffolds) were scattered among the initial groups. Based on the observations, we modeled the data with a mixture of two-dimensional Gaussian distributions

$$P(x_i) = \sum_{k=1}^G \tau_k \phi_k(x_i | \mu_k, \Sigma_k) \quad i = 1, 2, \dots, N,$$

where G is the number of components, τ_k is the probability that an observation belongs to the k th component ($\tau_k \geq 0$; $\sum_{k=1}^G \tau_k = 1$), x_i represents the observation in the data, N is the total number of scaffolds, and

$$\phi_k(x_i | \mu_k, \Sigma_k) = (2\pi)^{-\frac{p}{2}} |\Sigma_k|^{-\frac{1}{2}} \exp \left\{ -\frac{1}{2} (x_i - \mu_k)^T \Sigma_k^{-1} (x_i - \mu_k) \right\}.$$

The number of G components was set to the number of initial groups plus one, which was added to accommodate data points scattered among groups. The EM algorithm starting with M -step implemented in *mclust* package (Fraley and Raftery, 2002; Fraley et al., 2012) in R (Team, 2008) was applied to fit the mixed model. Here, each scaffold was weighted by length, thus the complete data log-likelihood is

$$l_c = \sum_{i=1}^n \sum_{k=1}^G z_{ik} w_i \log [\tau_k \phi_k(x_i | \mu_k, \Sigma_k)],$$

where n is the number of data points, z_{ik} is the conditional probability that observation i belongs to group k , w_i is the weight of observation i , in this case $w_i = l_i / \sum_{i=1}^n l_i$, and l_i is the length of scaffold i . The algorithm requires initial estimates of conditional probabilities. To simplify, we set the $z_{ik} = 1$, if observation i belonged to group k in the initial binning and to 0 if it did not.

Population Genome Assembly Validation

Owing to the limited reference genome sequences, it is still hard to investigate the completeness of each group from metagenomics data. Generally, different types of conserved genes have been chosen to validate the integrity of the genomic fragments, such as ribosomal genes and core genes within related organisms (Hess et al., 2011). However, ribosomal genes are not sufficient to assess the real completeness and accuracy of a metagenomics assembly, and the core genes are difficult to define. In this study, we evaluated the completeness of each group by 107 essential single-copy genes, which were conserved in 95% of all sequenced bacteria (Dupont et al., 2012). We referred to the steps described in Albertsen et al. (2013). First, we predicted the open reading frames (ORFs) of each group by Prodigal (Hyatt et al., 2012), these ORFs were then searched against a set of 107 hidden Markov models (HMMs) of essential single-copy genes by HMMER3 (Johnson et al., 2010) with the default settings.

Taxonomic and Functional Classifications

To understand the composition of this community, the predicted ORFs of each group are aligned to the NCBI-NR database using

BLASTP with a maximum e -value of $1e-5$. The blast outputs were then filtered (at least 40% amino acid sequence identity and 50% length hit) and the filtered results (each line was repeated by its FPKM) were imported into MEGAN5 (Huson et al., 2011) for taxonomic classification. To be more exact, the 16S ribosomal RNA genes of each group were predicted by RNAmmer (Lagesen et al., 2007) and the 16S ribosomal RNA genes of their closest neighbors were downloaded from the NCBI, the alignments were run in MEGA5 (Tamura et al., 2011) (neighbor-joining) at the default settings with 1 000 bootstraps. The predicted ORFs of each group were annotated using the GhostKOALA service (Kanehisa et al., 2012), and then the interesting KEGG pathways were analyzed comparatively. For the KEGG classification, the complementation ratio (the number of mapped genes divided by the total number of genes involved in the respective pathways) was counted (Endo et al., 2014). COGs (Clusters of Orthologous Groups) were also used to assign function (Tatusov et al., 2000).

RESULTS

Analysis of Metagenomics Data from *Microcystis* Colonies with an Improved Binning Method

Annual massive *Microcystis* blooms have become a severe ecological problem for Lake Taihu, the second largest lake in China. There are several species of *Microcystis* in the blooms and they have different colony morphology. The species used in this was *M. wesenbergii* T100 isolated from Lake Taihu. The culture was maintained in the lab for 8 years and the colony morphology was unchanged during the period of lab culture. For metagenomics study, one paired-end (PE) library with a 300-bp insert and one mate-pair (MP) library with a long insert (3 kb) were sequenced with 81 base PE reads, producing 6.6 and 3.2 Gb of raw data, respectively. After trimming and filtering, a total of 84 million clean reads were assembled into 60 Mb of scaffolds ranging in size from 1 kb to 1.2 Mb using SPAdes.

For the metagenomic datasets obtained from *Microcystis*-bacteria colonies, we developed an improved binning method, employing EM algorithm (Dempster et al., 1977) to assign individual scaffolds to each genome. We first used two genomic features, GC content and read coverage of each scaffold, as the two coordinates to generate a two-dimensional scatter-plot after DNA sequencing and assembly. The 2-D plot separated scaffolds according to the intrinsic properties of each genome. However, since the scaffolds with similar GC content and read coverage would tend to group together in the plot, further separation was needed. The values of GC content of scaffolds from a bacterial genome would fluctuate around the GC content of the whole genome. Statistically, a long scaffold would have a GC value closer to the GC value of the genome than a short scaffold and it is more reliable to assign long scaffolds to real genomic fragments of individual species than short ones or misassembled sequences. We therefore introduced scaffold length as another parameter so that each scaffold was weighted by its length and

further separation of scaffolds on the 2-D plot could be achieved. The number and position of population genome bins on the graphic can be determined based on the contour plot of the joint probability density function which is generated by kernel density estimation (KDE) (Rosenblatt, 1956; Parzen, 1962). We manually determined the value of contour level for binning so that the individual bin position can be separated in the best way from the others.

We tested this binning method first with a published simulated dataset in the article (Wu et al., 2014), which was generated from 10 species by MetaSim (Richter et al., 2008). The binning results showed that the assembly was regrouped into 10 genome bins (representing 10 genomes), and the majority of each bin was correctly assigned (Supplementary Figure S1). The performance of our binning method is comparable with MaxBin's (Wu et al., 2014), particularly for the species with high abundance.

The sequencing results of *M. wesenbergii* colonies mentioned previously were analyzed with our binning method. Based on the contour-plot results (Supplementary Figure S2), we applied a Gaussian mixture model to obtain the optimal scaffold clusters. The final binning result was visualized by a scatterplot with different colors representing different population genome bins (Figure 2). In total, nine genome bins (Supplementary Table 1) were identified and the population scaffolds with a coverage of 70.88% of all clean reads. The basic information of the groups (e.g., length, GC content, and the numbers in each group) is shown in Table 1. To identify these groups taxonomically, we analyzed them by the program MEGAN5 (Supplementary Figure S3). At the phylum level, six of the nine groups were Proteobacteria, which are the most abundant in Lake Taihu (Steffen et al., 2012), two were Bacteroidetes and one was the group of Cyanobacteria. At the order level, eight of the nine groups were classified to the relevant order with maximum probability, except for Group 3, which could contain more than

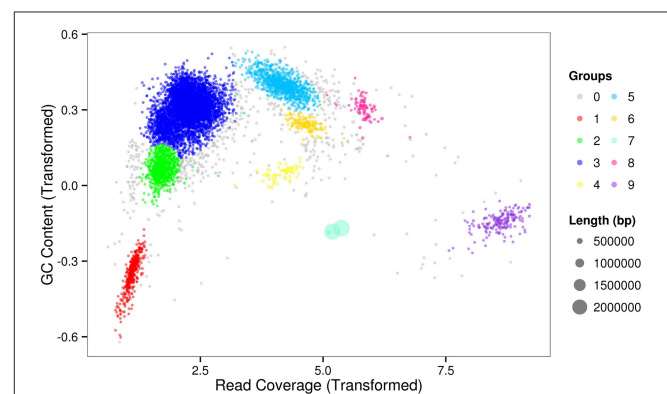
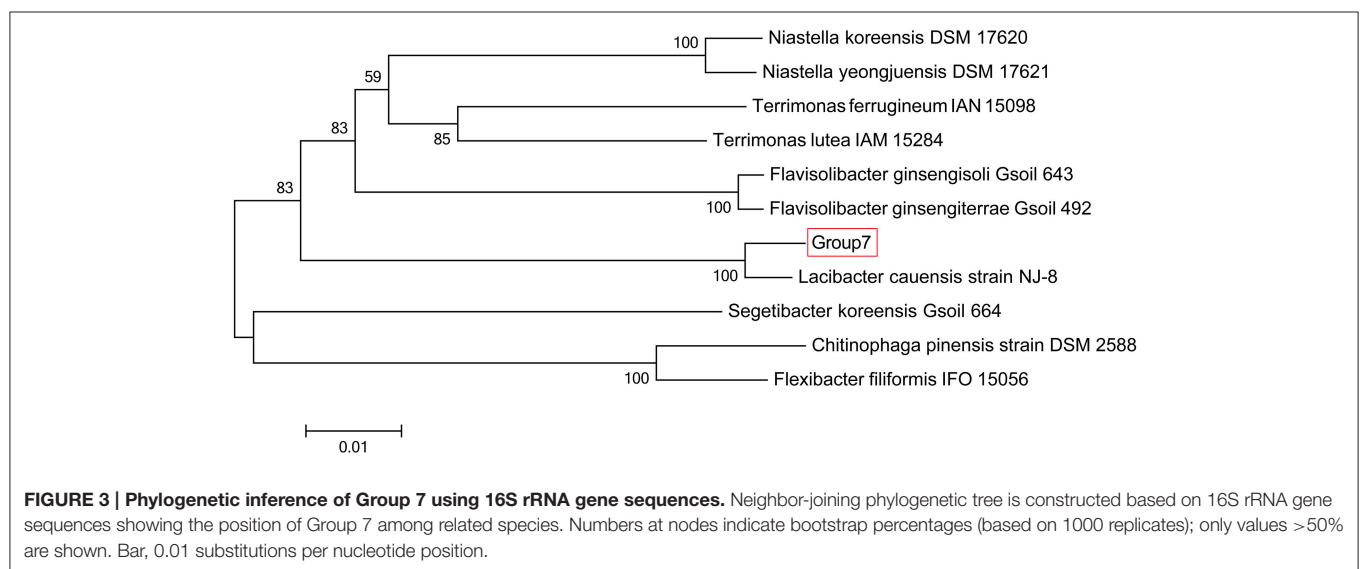


FIGURE 2 | Scatterplot of the final binning results for the metagenomics sample based on the Gaussian mixture model. The points in the figure represent scaffolds assembled from the microbial community and the point size indicates the length of scaffolds. The colored clusters on the graph are potential genomes of each group. The gray points dispersed around the clusters are scaffolds that were difficult for the model to assign.

TABLE 1 | The basic statistics for each group of *Microcystis* colonies.

Group	No. contigs	Length (bp)	GC content	No. essential genes	Relative abundance* (%)	Genus	Phylogenetic affiliation (phylum)
Group1	443	3,239,598	33.95	100/105(95.24)	0.5	<i>Flavobacterium</i>	Bacteroidetes
Group2	943	2,750,644	53.05	46/106(43.4)	0.5	<i>Limnobacter</i>	Betaproteobacteria
Group3	5810	10,026,678	64.26	91/105(86.67)	1.8	<i>Rhizobium</i>	Alphaproteobacteria
Group4	53	3,406,030	52.26	106/106(100)	4.4	<i>Limnobacter</i>	Betaproteobacteria
Group5	1000	4,081,424	68.5	93/105(88.57)	2.3	<i>Methylobacterium</i>	Alphaproteobacteria
Group6	150	4,587,864	61.67	104/105(99.05)	4.9	<i>Agrobacterium</i>	Alphaproteobacteria
Group7	3	4,737,723	41.25	105/105(100)	14.8	<i>Lacibacter</i>	Bacteroidetes
Group8	73	4,214,643	64.34	105/105(100)	6.4	<i>Pseudomonas</i>	Gammaproteobacteria
Group9	163	4,195,911	43.06	95/106(89.62)	31.6	<i>Microcystis</i>	Cyanobacteria

*Relative abundance was calculated as the percentage of reads of a genome bin in the total number of reads.



one order with similar GC content and abundance. At the genus level, Group 6, 7, 8, and 9 were from *Agrobacterium*, *Niastella*, *Pseudomonas*, and *Microcystis*, respectively. Group 1 and 5 were from genera that are closely related to *Flavobacterium* and *Methylobacterium*. Group 2 and 4 were both from *Limnobacter* and they had similar GC contents. Group 7 was classified to *Niastella* genus according to the result of MEGAN, but when the confirmation was performed based on the 16S ribosomal RNA, phylogenetic analysis showed that it was closer to *Lacibacter* genus (Figure 3). And one bacterium in the *Lacibacter* genus was isolated from the surface layer sediment of Lake Taihu and identified as *Lacibacter cauensis* NJ-8^T (Qu et al., 2009). To date, there is no reference genome for the *Lacibacter* genus, and our method provides near-complete chromosomes for this genus. This conflict between taxonomy assignments was due to the limitation of reference genomes in the analysis of metagenomics. The genomes of eight of the nine groups were estimated to be over 85% completed when compared with the number of essential genes of all sequenced bacteria at the phylum level (Dupont et al., 2012). Among the eight, five were over 95% complete (Table 1). The copy numbers of

the essential genes, which indicate how well the genomes are separated from one another, were also examined and the results (Supplementary Table 2) suggested that each group except Group 3 had little contamination from other genomes. As mentioned above, Group 3 could contain more than one bacterium since the number of its essential genes was much higher. Overall, our method effectively separated groups with different GC contents and abundances, while maintaining the integrity of each group.

Functional Analysis of the *Microcystis*-Dominated Community

To understand the biological functions of the microbial community, genes from all groups were assigned to broad functional categories (COGs) (Supplementary Figure S4). With the exception of general function prediction only (R) and function unknown (S), most of the genes were involved in amino acid transport and metabolism (E), cell wall/membrane/envelope biogenesis (M) and translation, ribosomal structure and biogenesis (J), inorganic ion transport and metabolism (P), and energy production and conversion (C). The result shows that

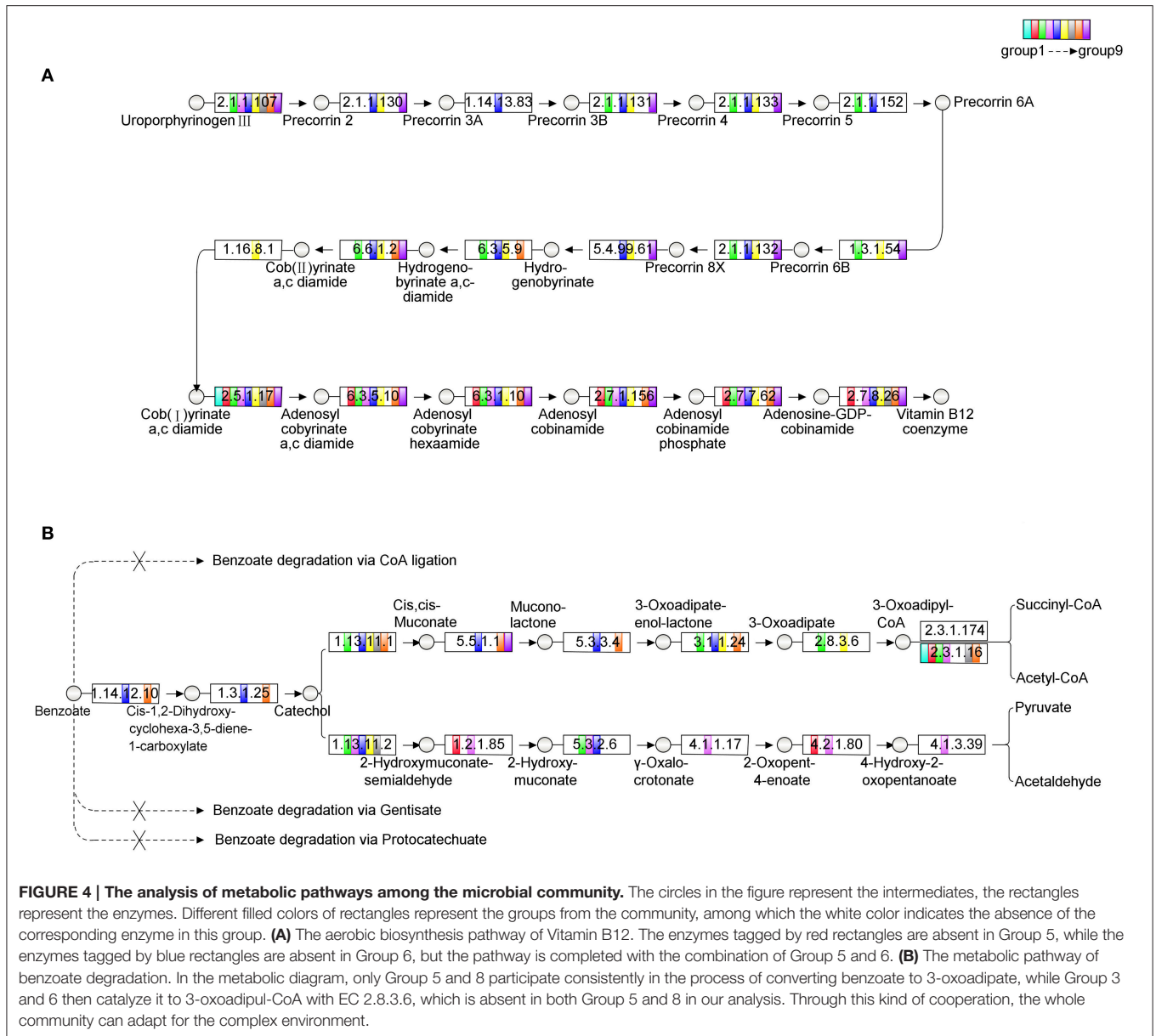


FIGURE 4 | The analysis of metabolic pathways among the microbial community. The circles in the figure represent the intermediates, the rectangles represent the enzymes. Different filled colors of rectangles represent the groups from the community, among which the white color indicates the absence of the corresponding enzyme in this group. **(A)** The aerobic biosynthesis pathway of Vitamin B12. The enzymes tagged by red rectangles are absent in Group 5, while the enzymes tagged by blue rectangles are absent in Group 6, but the pathway is completed with the combination of Group 5 and 6. **(B)** The metabolic pathway of benzoate degradation. In the metabolic diagram, only Group 5 and 8 participate consistently in the process of converting benzoate to 3-oxoadipate, while Group 3 and 6 then catalyze it to 3-oxoadipyl-CoA with EC 2.8.3.6, which is absent in both Group 5 and 8 in our analysis. Through this kind of cooperation, the whole community can adapt for the complex environment.

the quantity of each category in the Group 3 was three to four times more than that from other groups, suggesting this group contained more than one bacterium. This suggestion is supported by the result of taxonomic assignment (Table 1).

Microcystis is the only photoautotrophic bacterium and the other bacteria are all heterotrophic in the *Microcystis*-bacteria colonies (Cole, 1982). Since the culture was maintained with a minimal medium, these heterotrophic bacteria depended on both carbon and energy source from *Microcystis*. On the other hand, analysis of *Microcystis* genomes showed that it had a methionine biosynthesis pathway that has a type-II MetH enzyme which needs vitamin B12 as cofactor (Croft et al., 2005; Kaneko et al., 2008; Helliwell et al., 2011). During the subculture, no vitamin B12 was added to culture. The possible explanation of the fact was that these heterotrophic bacteria are the likely candidates that

providing vitamin B12 for *Microcystis*. An alternative was that these bacteria could provide methionine directly to *Microcystis* cells. In either case, *Microcystis* benefited from being associated with these bacteria. Therefore, the relationship between the heterotrophic bacteria and *Microcystis* is mutually beneficial. The contribution of each individual bacterium in the colonies to the biosynthesis of vitamin B12 was analyzed and the results showed that none of group could complete the aerobic biosynthesis pathway of vitamin B12 independently. The complementation ratio of the vitamin B12 biosynthesis pathway of the colonies (Supplementary Figure S5) calculated according to Endo et al. (2014) suggested that two groups, such as Group 5 and Group 6 (Figure 4A), or more of the bacteria were required for the vitamin B12 synthesis in the colonies. The complementation ratio of benzoate degradation pathway was also indicative of

mutualism. Group 5 and 8 could be responsible for converting benzoate to 3-oxoadipate (Harayama et al., 1991), Group 3 and 6 then could catalyze it to 3-oxoadipul-CoA with EC 2.8.3.6, which is absent from Group 5 and 8 (**Figure 4B**). While the cellular mechanisms of the metabolic complementation require more studies, our analysis implied that benzoate can be degraded to acetyl-CoA and succinyl-CoA through cooperated actions of these species.

The harmful effects of *Microcystis* blooms to human and animal health are noteworthy, including the release of toxic microcystins and lipopolysaccharides (LPS). A typical LPS molecule is composed of three parts: lipid A, core polysaccharides, and O-antigen repeats. Lipid A is located in the outer leaflet of the outer membrane, while core polysaccharides and O-antigen repeats lie on the surface of the bacterial cells by attaching to lipid A. Although *Microcystis* LPS have been reported to cause some serious health issues, their synthesis are not well understood (Raziuddin et al., 1983; Snyder et al., 2009). We analyzed the LPS biosynthesis pathways (Supplementary Figure S6) and found that *Microcystis* only had genes for synthesis of lipid A and lacked the genes responsible for synthesis of other components of LPS. Figure S6 also shows that the associated bacteria had almost complete composition of LPS, suggesting that the LPS-related health problems were caused by the *Microcystis*-bacterial colonies.

DISCUSSION

Colony formation is one of the key features in *Microcystis* blooms and it has been shown that heterotrophic bacteria play important roles in the colony formation (Shen et al., 2011; Wang et al., 2015). The *M. wesenbergii* T100 was isolated from Lake Taihu and it maintained the ability to form colonies in subculture. The lab maintained colonies are less complex and their bacterial composition is less variable than the ones in nature and therefore they are more suitable for our initial metagenomics study of colony structures.

In analysis of metagenomic data, sequence binning is one of the most important steps for obtaining more information about composition and structure the microbial community. Comparing with single bacteria genome assembly, metagenomic assembly is more fragmental and it generates more small scaffolds. How to assign small scaffolds, which have been often treated as noise, to a respective genome is a challenge for metagenomic binning. The *N50* values (a statistic of a set of contig or scaffold lengths, widely used in genome assembly) of individual population genome bins in metagenomic assembly vary in a wide range. The scaffold number of each genomics bin varies from one to thousands, causing various problems for model-based binning. Our binning method employed the EM algorithm (Dempster et al., 1977) to assign the assembled scaffolds, which were derived from individual genomes, to their respective groups. The parameters used in our method are GC content, contig coverage and contig length. GC content is one of the most intrinsic compositional features of genomes and it is often used in the identification of unknown DNA fragments (Karlin, 1998) and contig/scaffold coverage level has been used in other binning

methods (Wrighton et al., 2012; Albertsen et al., 2013). A big difference between our binning approach and other methods based on mixture model is that the joint distribution of sequence features, GC content and abundance, was dissected at nucleotide level instead of sequence level and it was accomplished by introduction of contig length as the third parameter. The finer granularity could provide higher resolution and reduce the noise of poor-assembled short scaffolds. With accurate initial values of parameters obtained by kernel density estimation (KDE) (Rosenblatt, 1956; Parzen, 1962), the number of iteration in calculation could be kept minimal. It brought a big improvement in binning and made the following EM algorithm calculation more reliable.

Our metagenomics study of the *Microcystis* colonies showed that various species were associated with the colonies following subculturing. While the dominant organism in the *Microcystis* colonies was *Microcystis wesenbergii* as expected, at least 8 other microbes were found. Analysis of the published metagenomic samples from Lake Taihu showed that these bacteria were present in natural environments (Li et al., 2011; Steffen et al., 2012), suggesting that they were associated with *Microcystis* colonies before the colonies were isolated (Supplementary Table 3). For example, Li et al. (2011) studied the microbial and functional diversity of a bloom in Lake Taihu by Roche 454 platform, eight groups (bacteria species) associated with *Microcystis* in our results were consistent with the taxonomic groups detected by the 454 platform. Our results were also in agreement with the taxonomy of heterotrophic bacteria isolated from *Microcystis* colonies (Shen et al., 2011). Another example is that Group 8 was assigned to *Pseudomonas*, which is a common genus in the phycosphere (Wu et al., 2007; Berg et al., 2008; Li et al., 2011).

A mutually beneficial relationship among the bacteria in the colonies is observed. The energy and carbon flows among the bacteria in the colonies are predictable since *Microcystis* was the only photoautotrophic species in the colonies and the other heterotrophic bacteria depend on *Microcystis* for growth under the culture conditions. On the other hand, *Microcystis* was dependent upon the heterotrophic bacteria for vitamin B-12 that is required for its growth. Analysis of metabolic pathways based on genome information suggested that more than one member of heterotrophic bacteria was involved in biosynthesis of vitamin B12 in the colonies (**Figure 4A**).

Besides *M. wesenbergii* T100, we have obtained other *Microcystis* from Lake Taihu, including *M. aeruginosa*, and they can be kept forming colonies in the lab. However, when a strain of *M. aeruginosa* was isolated (Yang et al., 2013, 2015) and kept in a bacterium-free condition, it produced much less extracellular polysaccharides and lost the ability to form colony. This observation and the reports that bacteria play a role in colony formation (Shen et al., 2011; Wang et al., 2015) led us speculate that extracellular polysaccharides of *Microcystis* colonies could be synthesized by more than one member of the colony. This view is supported by our analysis of synthesis of lipopolysaccharides (LPS) in the colonies. Among harmful substances produced by *Microcystis* blooms, LPS are classified as endotoxin (Raetz et al., 2007; Wang and Quinn, 2010) and can cause severe diseases in humans and animals (Stewart et al., 2006;

Kusumoto et al., 2010). Our metabolic analysis showed that it was unlikely that the *Microcystis* endotoxins were produced by *Microcystis* alone since it didn't contain all the genes required for LPS biosynthesis and one or more members of the *Microcystis*-bacteria community are involved in the process (Best et al., 2003; Bernardová et al., 2008).

Our study demonstrated a mutually beneficial and inter-dependent relationship among the bacteria in *Microcystis* colonies maintained in a subculture condition. It should be realized that the *Microcystis* blooms are more complex and there might be some bacteria that disassociated from the colonies after the long subculture in the lab. It is also worthwhile to note that while our binning method is well suitable for datasets derived from microbial communities with stable species composition (up to 30 species based on the test of published data) (Supplementary Figure S7, Supplementary Table 4), the resolving power of this method could be limited in situations where the species in a community are closely related in phylogeny, or a large number of species is present in a community and co-assembly problem could not be prevented. Nevertheless, the metagenomic approach reported in this study and the results of metabolic pathway complementation could serve as a basis for future study of a more complex interaction between *Microcystis* and the associated bacteria in natural environment.

CONCLUSION

Here, we presented a visualization-enhanced binning method and applied it to analyze cyanobacteria-dominated bloom communities from Lake Taihu, China, reconstructing individual bacterial genomes from metagenomic assembly. By analyzing the metabolic pathways of the microbial community, cooperative interactions among the complex species were indicated, which provided insight into the formation mechanism of cyanobacterial blooms.

REFERENCES

- Albertsen, M., Hugenholtz, P., Skarshewski, A., Nielsen, K. L., Tyson, G. W., and Nielsen, P. H. (2013). Genome sequences of rare, uncultured bacteria obtained by differential coverage binning of multiple metagenomes. *Nat. Biotechnol.* 31, 533–538. doi: 10.1038/nbt.2579
- Alneberg, J., Bjarnason, B. S., Bruijn, I. D., Schirmer, M., Quick, J., Ijaz, U. Z., et al. (2014). Binning metagenomic contigs by coverage and composition. *Nat. Methods* 11, 1144–1146. doi: 10.1038/nmeth.3103
- Bankevich, A., Nurk, S., Antipov, D., Gurevich, A. A., Dvorkin, M., Kulikov, A. S., et al. (2012). SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. *J. Comput. Biol.* 19, 455–477. doi: 10.1089/cmb.2012.0021
- Berg, K. A., Lyra, C., Sivonen, K., Paulin, L., Suomalainen, S., Tuomi, P., et al. (2008). High diversity of cultivable heterotrophic bacteria in association with cyanobacterial water blooms. *ISME J.* 3, 314–325. doi: 10.1038/ismej.2008.110
- Bernardová, K., Babica, P., Maršálek, B., and Bláha, L. (2008). Isolation and endotoxin activities of lipopolysaccharides from cyanobacterial cultures and complex water blooms and comparison with the effects of heterotrophic bacteria and green alga. *J. Appl. Toxicol.* 28, 72–77. doi: 10.1002/jat.1257
- Best, J. H., Eddy, F. B., and Codd, G. A. (2003). Effects of *Microcystis* cells, cell extracts and lipopolysaccharide on drinking and liver function in rainbow

ACCESSION CODES

The raw sequences data reported in this study have been submitted to the National Center for Biotechnology Information Sequence Read Archive under accession numbers SRX993730 and SRX1007860. This Whole Genome Shotgun project has been deposited at DDBJ/EMBL/GenBank under the accession LFEF00000000. The version described in this paper is version LFEF01000000.

AUTHOR CONTRIBUTIONS

TL and ZL conceived the study and designed the experiments. ZL, MX, and MR carried out the analysis of data. TL, MX, and MR prepared the first draft of the manuscript. CY contributed materials. HY contributed to the data analysis. ZL, JZ, and TL revised and polished the manuscript. All authors participated in the discussion of manuscript and have agreed to the final content.

ACKNOWLEDGMENTS

We thanked Dr. Renhui Li for providing the sample. The study was supported by the National Natural Science Foundation of China (Grant no. 31000113), the National High-Technology Research and Development Program of China (863 Program, 2009AA02Z30) and the Autonomous Projects of the State Key Laboratory of Freshwater Ecology and Biotechnology for funding (2011FBZ31 and 2011FBZ32).

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <http://journal.frontiersin.org/article/10.3389/fmicb.2016.00056>

- trout *Oncorhynchus mykiss* Walbaum. *Aquat. Toxicol.* 64, 419–426. doi: 10.1016/S0166-445X(03)00105-X
- Bláha, L., Babica, P., and Maršálek, B. (2009). Toxins produced in cyanobacterial water blooms - toxicity and risks. *Interdiscip. Toxicol.* 2, 36–41. doi: 10.2478/v10102-009-0006-2
- Bolger, A. M., Lohse, M., and Usadel, B. (2014). Trimmomatic: a flexible trimmer for illumina sequence data. *Bioinformatics* 30, 2114–2120. doi: 10.1093/bioinformatics/btu170
- Cole, J. J. (1982). Interactions between bacteria and algae in aquatic ecosystems. *Ann. Rev. Ecol. Evol. Syst.* 13, 291–314. doi: 10.1146/annurev.es.13.110182.001451
- Conley, D. J., Paerl, H. W., Howarth, R. W., Boesch, D. F., Seitzinger, S. P., Havens, K. E., et al. (2009). Controlling eutrophication: nitrogen and phosphorus. *Science* 323, 1014–1015. doi: 10.1126/science.1167755
- Croft, M. T., Lawrence, A. D., Raux-Deery, E., Warren, M. J., and Smith, A. G. (2005). Algae acquire vitamin B12 through a symbiotic relationship with bacteria. *Nature* 438, 90–93. doi: 10.1038/nature04056
- Davis, T. W., Berry, D. L., Boyer, G. L., and Gobler, C. J. (2009). The effects of temperature and nutrients on the growth and dynamics of toxic and non-toxic strains of *Microcystis* during cyanobacteria blooms. *Harmful Algae* 8, 715–725. doi: 10.1016/j.hal.2009.02.004

- Dempster, A. P., Laird, N. M., and Rubin, D. B. (1977). Maximum likelihood from incomplete data via the EM algorithm. *J. R. Stat. Soc.* 39, 1–38.
- Dick, G. J., Andersson, A. F., Baker, B. J., Simmons, S. L., Thomas, B. C., Yelton, A. P., et al. (2009). Community-wide analysis of microbial genome sequence signatures. *Genome Biol.* 10, R85. doi: 10.1186/gb-2009-10-8-r85
- Dupont, C. L., Rusch, D. B., Yooseph, S., Lombardo, M.-J., Alexander Richter, R., Valas, R., et al. (2012). Genomic insights to SAR86, an abundant and uncultivated marine bacterial lineage. *ISME J.* 6, 1186–1199. doi: 10.1038/ismej.2011.189
- Dziallas, C., and Grossart, H.-P. (2011). Temperature and biotic factors influence bacterial communities associated with the cyanobacterium *Microcystis* sp. *Environ. Microbiol.* 13, 1632–1641. doi: 10.1111/j.1462-2920.2011.02479.x
- Dziallas, C., and Grossart, H.-P. (2012). Microbial interactions with the cyanobacterium *Microcystis aeruginosa* and their dependence on temperature. *Mar. Biol.* 159, 2389–2398. doi: 10.1007/s00227-012-1927-4
- Eiler, A., and Bertilsson, S. (2004). Composition of freshwater bacterial communities associated with cyanobacterial blooms in four Swedish lakes. *Environ. Microbiol.* 6, 1228–1243. doi: 10.1111/j.1462-2920.2004.00657.x
- Eiler, A., Heinrich, F., and Bertilsson, S. (2011). Coherent dynamics and association networks among lake bacterioplankton taxa. *ISME J.* 6, 330–342. doi: 10.1038/ismej.2011.113
- Endo, A., Watanabe, T., Ogata, N., Nozawa, T., Aikawa, C., Arakawa, S., et al. (2014). Comparative genome analysis and identification of competitive and cooperative interactions in a polymicrobial disease. *ISME J.* 9, 629–642. doi: 10.1038/ismej.2014.155
- Falconer, I. R., and Humpage, A. R. (2006). Cyanobacterial (Blue-Green Algal) toxins in water supplies: cylindrospermopsins. *Environ. Toxicol.* 21, 299–304. doi: 10.1002/tox.20194
- Fraley, C., and Raftery, A. (2002). Model-based clustering, discriminant analysis, and density estimation. *J. Am. Stat. Assoc.* 97, 611–631. doi: 10.1198/016214502760047131
- Fraley, C., Raftery, A., Murphy, T., and Scrucca, L. (2012). *mclust Version 4 for R: Normal Mixture Modeling for Model-Based Clustering, Classification, and Density Estimation*. Technical Report.
- Harayama, S., Reik, M., Bairoch, A., Neidle, E. L., and Ornston, L. N. (1991). Potential DNA slippage structures acquired during evolutionary divergence of *Acinetobacter calcoaceticus* chromosomal benABC and *Pseudomonas putida* TOL pWW0 plasmid xylXYZ, genes encoding benzoate dioxygenases. *J. Bacteriol.* 173, 7540–7548.
- Helliwell, K. E., Wheeler, G. L., Leptos, K. C., Goldstein, R. E., and Smith, A. G. (2011). Insights into the evolution of vitamin B12 auxotrophy from sequenced algal genomes. *Mol. Biol. Evol.* 28, 2921–2933. doi: 10.1093/molbev/msr124
- Hess, M., Sczyrba, A., Egan, R., Kim, T.-W., Chokhwal, H., Schroth, G., et al. (2011). Metagenomic discovery of biomass-degrading genes and genomes from cow rumen. *Science* 331, 463–467. doi: 10.1126/science.1200387
- Huson, D. H., Auch, A. F., Qi, J., and Schuster, S. C. (2007). MEGAN analysis of metagenomic data. *Genome Res.* 17, 377–386. doi: 10.1101/gr.5969107
- Huson, D. H., Mitra, S., Ruscheweyh, H. J., Weber, N., and Schuster, S. C. (2011). Integrative analysis of environmental sequences using MEGAN4. *Genome Res.* 21, 1552–1560. doi: 10.1101/gr.120618.111
- Hyatt, D., LoCascio, P. F., Hauser, L. J., and Uberbacher, E. (2012). Gene and translation initiation site prediction in metagenomic sequences. *Bioinformatics* 28, 2223–2230. doi: 10.1093/bioinformatics/bts429
- Johnson, L. S., Eddy, S. R., and Portugaly, E. (2010). Hidden Markov model speed heuristic and iterative HMM search procedure. *BMC Bioinformatics* 11:431. doi: 10.1186/1471-2105-11-431
- Kanehisa, M., Goto, S., Sato, Y., Furumichi, M., and Tanabe, M. (2012). KEGG for integration and interpretation of large-scale molecular data sets. *Nucleic Acids Res.* 40, 109–114. doi: 10.1093/nar/gkr988
- Kaneko, T., Nakajima, N., Okamoto, S., Suzuki, I., Tanabe, Y., Tamaoki, M., et al. (2008). Complete genomic structure of the bloom-forming toxic cyanobacterium *Microcystis aeruginosa* NIES-843. *DNA Res.* 14, 247–256. doi: 10.1093/dnares/dsm026
- Karlin, S. (1998). Global dinucleotide signatures and analysis of genomic heterogeneity. *Microbiology* 1, 598–610. doi: 10.1016/s1369-5274(98)80095-7
- Kunin, V., Copeland, A., Lapidus, A., Mavromatis, K., and Hugenholtz, P. (2008). A bioinformatician's guide to metagenomics. *Microbiol. Mol. Biol. Rev.* 72, 557–578. doi: 10.1128/MMBR.00009-08
- Kusumoto, S., Fukase, K., and Shiba, T. (2010). Key structures of bacterial peptidoglycan and lipopolysaccharide triggering the innate immune system of higher animals: chemical synthesis and functional studies. *Proc. Jpn. Acad. B* 86, 322–337. doi: 10.2183/pjab.86.322
- Laczny, C. C., Pinel, N., Vlassis, N., and Wilmes, P. (2014). Alignment-free visualization of metagenomic data by nonlinear dimension reduction. *Sci. Rep.* 4, 4516–4528. doi: 10.1038/srep04516
- Lagesen, K., Hallin, P., Rodland, E. A., Staerfeldt, H. H., Rognes, T., and Ussery, D. W. (2007). RNAmmmer: consistent and rapid annotation of ribosomal RNA genes. *Nucleic Acids Res.* 35, 3100–3108. doi: 10.1093/nar/gkm160
- Langmead, B., Trapnell, C., Pop, M., and Salzberg, S. L. (2009). Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol.* 10, R25. doi: 10.1186/gb-2009-10-3-r25
- Li, N., Zhang, L., Li, F., Wang, Y., Zhu, Y., Kang, H., et al. (2011). Metagenome of microorganisms associated with the toxic Cyanobacteria *Microcystis aeruginosa* analyzed using the 454 sequencing platform. *Chin. J. Oceanol. Limnol.* 29, 505–513. doi: 10.1007/s00343-011-0056-0
- Li, R., Carmichael, W. W., Brittain, S., Eaglesham, G. K., Shaw, G. R., Mahakant, A., et al. (2001). Isolation and identification of the cyanotoxin cylindrospermopsin and deoxy-cylindrospermopsin from a Thailand strain of *Cylindrospermopsis raciborskii* (Cyanobacteria). *Toxicon* 39, 973–980. doi: 10.1016/S0041-0101(00)00236-1
- Ma, J., Brookes, J. D., Qin, B., Paerl, H. W., Gao, G., Wu, P., et al. (2014). Environmental factors controlling colony formation in blooms of the cyanobacteria *Microcystis* spp. in Lake Taihu, China. *Harmful Algae* 31, 136–142. doi: 10.1016/j.hal.2013.10.016
- Mackelprang, R., Waldrop, M. P., DeAngelis, K. M., David, M. M., Chavarria, K. L., Blazewicz, S. J., et al. (2011). Metagenomic analysis of a permafrost microbial community reveals a rapid response to thaw. *Nature* 480, 368–371. doi: 10.1038/nature10576
- Mande, S. S., Mohammed, M. H., and Ghosh, T. S. (2012). Classification of metagenomic sequences: methods and challenges. *Brief. Bioinformatics* 13, 669–681. doi: 10.1093/bib/bbs054
- Mou, X., Jacob, J., Lu, X., Robbins, S., Sun, S., and Ortiz, J. D. (2013). Diversity and distribution of free-living and particle-associated bacterioplankton in Sandusky Bay and adjacent waters of Lake Erie Western Basin. *J. Great Lakes Res.* 39, 352–357. doi: 10.1016/j.jglr.2013.03.014
- Pachter, L. (2011). Models for transcript quantification from RNA-Seq. arXiv preprint arXiv, 11043889.
- Parzen, E. (1962). On estimation of a probability density function and mode. *Ann. Math. Stat.* 33, 1065–1076. doi: 10.1214/aoms/1177704472
- Pope, P. B., and Patel, B. K. (2008). Metagenomic analysis of a freshwater toxic cyanobacteria bloom. *FEMS Microbiol. Ecol.* 64, 9–27. doi: 10.1111/j.1574-6941.2008.00448.x
- Qu, J. H., Yuan, H. L., Yang, J. S., Li, H. F., and Chen, N. (2009). *Lacibacter cauensis* gen. nov., sp. nov., a novel member of the phylum Bacteroidetes isolated from sediment of a eutrophic lake. *Int. J. Syst. Evol. Microbiol.* 59, 1153–1157. doi: 10.1099/ijs.0.003475-0
- Raetz, C. R., Reynolds, C. M., Trent, M. S., and Bishop, R. E. (2007). Lipid a modification systems in gram-negative bacteria. *Annu. Rev. Biochem.* 76, 295–329. doi: 10.1146/annurev.biochem.76.010307.145803
- Raziuddin, S., Siegelman, H., and Tornabene, T. (1983). Lipopolysaccharides of the cyanobacterium *Microcystis aeruginosa*. *Eur. J. Biochem.* 137, 333–336. doi: 10.1111/j.1432-1033.1983.tb07833.x
- Richter, D. C., Ott, F., Auch, A. F., Schmid, R., and Huson, D. H. (2008). MetaSim—a sequencing simulator for genomics and metagenomics. *PLoS ONE* 3:e3373. doi: 10.1371/journal.pone.0003373
- Roberts, A., and Pachter, L. (2013). Streaming fragment assignment for real-time analysis of sequencing experiments. *Nat. Methods* 10, 71–73. doi: 10.1038/nmeth.2251
- Rosenblatt, M. (1956). Remarks on some nonparametric estimates of a density function. *Ann. Math. Stat.* 27, 832–837. doi: 10.1214/aoms/1177728190
- Sheather, S. J., and Jones, M. C. (1991). A reliable data-based bandwidth selection method for kernel density estimation. *J. R. Stat. Soc. B* 53, 683–690.
- Shen, H., Niu, Y., Xie, P., Tao, M. I. N., and Yang, X. I. (2011). Morphological and physiological changes in *Microcystis aeruginosa* as a result of interactions with heterotrophic bacteria. *Freshw. Biol.* 56, 1065–1080. doi: 10.1111/j.1365-2427.2010.02551.x

- Sigee, D. C. (2005). *Freshwater Microbiology: Biodiversity and Dynamic Interactions of Microorganisms in the Freshwater Environment*. Chichester: John Wiley & Sons.
- Sivonen, K., Jones, G. (1999). *Cyanobacterial Toxins. Toxic Cyanobacteria in Water: A Guide to Public Health Significance, Monitoring and Management*. London: E & FN Spon.
- Snyder, D. S., Brahamsha, B., Azadi, P., and Palenik, B. (2009). Structure of compositionally simple lipopolysaccharide from marine *Synechococcus*. *J. Bacteriol.* 191, 5499–5509. doi: 10.1128/JB.00121-09
- Steffen, M. M., Li, Z., Effler, T. C., Hauser, L. J., Boyer, G. L., and Wilhelm, S. W. (2012). Comparative metagenomics of toxic freshwater cyanobacteria bloom communities on two continents. *PLoS ONE* 7:e44002. doi: 10.1371/journal.pone.0044002
- Stewart, I., Schluter, P. J., and Shaw, G. R. (2006). Cyanobacterial lipopolysaccharides and human health - a review. *Environ. Health* 5, 7–30. doi: 10.1186/1476-069X-5-7
- Strous, M., Kraft, B., Bisdorf, R., and Tegetmeyer, H. E. (2012). The binning of metagenomic contigs for microbial physiology of mixed cultures. *Front. Microbiol.* 3:410. doi: 10.3389/fmicb.2012.00410
- Tamura, K., Peterson, D., Peterson, N., Stecher, G., Nei, M., and Kumar, S. (2011). MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. *Mol. Biol. Evol.* 28, 2731–2739. doi: 10.1093/molbev/msr121
- Tatusov, R. L., Galperin, M. Y., Natale, D. A., and Koonin, E. V. (2000). The COG database: a tool for genome-scale analysis of protein functions and evolution. *Nucleic Acids Res.* 28, 33–36. doi: 10.1093/nar/28.1.33
- Team, R. C. (2008). *R: A Language and Environment for Statistical Computing*. Vienna: R Foundation for Statistical Computing, 1–1731.
- Trapnell, C., Williams, B. A., Pertea, G., Mortazavi, A., Kwan, G., and Baren, M. J. V., et al. (2010). Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation. *Nat. Biotechnol.* 28, 511–515. doi: 10.1038/nbt.1621
- Tyson, G., Chapman, J., Hugenholtz, P., Allen, E. E., Ram, R. J., Richardson, P. M., et al. (2004). Community structure and metabolism through reconstruction of microbial genomes from the environment. *Nature* 428, 37–43. doi: 10.1038/nature02340
- Venter, J. C., Remington, K., Heidelberg, J. F., Halpern, A. L., Rusch, D., Eisen, J. A., et al. (2004). Environmental genome shotgun sequencing of the sargasso sea. *Science* 304, 66–74. doi: 10.1126/science.1093857
- Wand, M. (1994). Fast computation of multivariate kernel estimators. *J. Comput. Graph. Stat.* 3, 433–445.
- Wand, M., and Jones, M. (1995). *Kernel Smoothing*, Vol. 60. London: Chapman and Hall/CRC. doi: 10.1007/978-1-4899-4493-1
- Wang, W., Shen, H., Shi, P., Chen, J., Ni, L., and Xie, P. (2015). Experimental evidence for the role of heterotrophic bacteria in the formation of Microcystis colonies. *J. Appl. Phycol.* doi: 10.1007/s10811-015-0659-5. [Epub ahead of print].
- Wang, X., and Quinn, P. J. (2010). Lipopolysaccharide: biosynthetic pathway and structure modification. *Prog. Lipid Res.* 49, 97–107. doi: 10.1016/j.plipres.2009.06.002
- Wang, Y., Leung, H. C. M., Yiu, S. M., and Chin, F. Y. L. (2012). MetaCluster 5.0: a two-round binning approach for metagenomic data for low-abundance species in a noisy sample. *Bioinformatics* 28, i356–i362. doi: 10.1093/bioinformatics/bts397
- Wrighton, K. C., Thomas, B. C., Sharon, I., Miller, C. S., Castelle, C. J., VerBerkmoes, N. C., et al. (2012). Fermentation, hydrogen, and sulfur metabolism in multiple uncultivated bacterial phyla. *Science* 337, 1661–1665. doi: 10.1126/science.1224041
- Wu, X., Xi, W., Ye, W., and Yang, H. (2007). Bacterial community composition of a shallow hypertrophic freshwater lake in China, revealed by 16S rRNA gene sequences. *FEMS Microbiol. Ecol.* 61, 85–96. doi: 10.1111/j.1574-6941.2007.00326.x
- Wu, Y.-W., Tang, Y.-H., Tringe, S. G., Simmons, B. A., and Singer, S. W. (2014). MaxBin: an automated binning method to recover individual genomes from metagenomes using an expectation-maximization algorithm. *Microbiome* 2, 26–44. doi: 10.1186/2049-2618-2-26
- Wu, Y.-W., and Ye, Y. (2011). A novel abundance-based algorithm for binning metagenomic sequences using l-tuples. *J. Comput. Biol.* 18, 523–534. doi: 10.1089/cmb.2010.0245
- Yang, C., Lin, F., Li, Q., Li, T., and Zhao, J. (2015). Comparative genomics reveals diversified CRISPR-Cas systems of globally distributed *Microcystis aeruginosa*, a freshwater bloom-forming cyanobacterium. *Front. Microbiol.* 6:394. doi: 10.3389/fmicb.2015.00394
- Yang, C., Zhang, W., Ren, M., Song, L., Li, T., and Zhao, J. (2013). Whole-genome sequence of *Microcystis aeruginosa* TAIHU98, a nontoxic bloom-forming strain isolated from Taihu Lake, China. *Genome Announc.* 1, e00333–e00313. doi: 10.1128/genomeA.00333-13

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2016 Xie, Ren, Yang, Yi, Li, Li and Zhao. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.