ANALYSIS OF THE MICROCYSTIN-LR PRODUCTION ABILITY OF METAGENOMIC *MCY* GENES IN FRESHWATER AQUACULTURE PONDS FOCUSING ON THE ABUNDANCE OF METAGENOMIC *MCY* GENES AND SNP

DONG, S. J.[1] – BI, X. D.[1*] – ZHANG, J. Y.[2] – DAI, W.[1*] – ZHANG, P. P.[3] – ZHOU, K. R.[3] – WANG, X. Y.[1] – ZHANG, D. J.[1]

[1]*Key Laboratory for Aqua-Ecology and Aquaculture of Tianjin, Department of Fisheries Science, Tianjin Agricultural University, Tianjin 300384, China*

[2]*Algae-Hub Big Data Center, MetaBio Science &Technology Co., Ltd., Wuxi 214135, China*

[3]*Jiangsu Wuxi Environmental Monitoring Centre, Wuxi 214023, China*

[*]*Corresponding authors*
e-mail: yl801123@aliyun.com, daiweitj@126.com

(Received 29th Jul 2020; accepted 19th Nov 2020)

Abstract. MCs pollution have become a worldwide problem for freshwater aquaculture, and MC-LR has attracted considerable attention due to its potent hepatotoxicity. To analysis the MC-LR production ability of cyanobacteria in field, the relationship between the MC-LR production ability and MC-LR content, TN content, TP content, SNP of the metagenomic *mcy* genes were investigated in 5 freshwater fishponds during July 2018. The results showed that MC-LR content was significantly positively correlated with the abundance of metagenomic *mcy* gene A-J, while the MC-LR production ability of metagenomic *mcy* gene A-J was in indistinctive correlation with the abundance of metagenomic mcy gene, and MC-LR content, and the content of TN and TP in sufficient nitrogen and phosphate conditions. 18 SNPs significantly positively correlated with the MC-LR production ability and 388 SNPs significantly negatively correlated with the MC-LR production ability. The genotype of *mcy* C and *mcy* B in *Microcystis aeruginosa* NIES-843 was the most favorable for the MC-LR production in present study. There might be some mechanism to avoid substrate specific changes caused by genetic mutation of *mcy* C.
Keywords: *harmful cyanobacteria, toxin, nutrient level, microcystin synthetase genes, correlation analysis*

Abbreviations: MCs: microcystins; MC-LR: microcystins-LR; TN: total nitrogen; TP: total phosphate; SNP: the single nucleotide polymorphism; *mcy*: the microcystin synthetase; NRPS: nonribosomal peptide synthetase; PKS: polyketide synthase; HPLC: high performance liquid chromatography; Adda: 3- amino-9-methoxy-2, 6, 8-trimethyl-10-phenyldeca-4, 6-dienoic acid; MeAsp: methylaspartic acid; Mdha: methyldehydroalanine

Introduction

In natural freshwater bodies, cyanobacteria genera, including *Anabaena*, *Aphanocapsa*, *Hapalosiphon*, *Nostoc*, *Oscillatoria*, *Pseudanabaena*, *Planktothrix* and *Microcystis*, often produce MCs (Falconer, 1999; Tillett et al., 2000; Carmichael, 2001; Codd et al., 2005). MCs are classified as possible human carcinogens (class B) by the International Agency for Research on Cancer (IARC) (Ngwa et al., 2014). MCs are a family of monocyclic heptapeptides hepatotoxins. They share a general structure of cyclic [-D-Ala1-L-X2-MeAsp3-L-Z4-Adda5-D-Glu6-N-Mdha7], in which X and Z represent two variable L-amino acids (Rastogi et al., 2014).

Microcystin-production cyanobacteria contain the *mcy* gene cluster. MCs cannot be produced when one or more of the required *mcy* genes are absent via gene deletion,

http://www.aloki.hu ● ISSN 1589 1623 (Print) ● ISSN1785 0037 (Online)
DOI: http://dx.doi.org/10.15666/aeer/1901_237248

recombination, and transformation, or disrupted and inactivated by transposons or phages (Ngwa et al., 2014; Zuo et al., 2018). The *mcy* gene cluster spanning 55kb includes NRPS genes, PKS genes, fused NRPS-PKS genes and modifier genes (Noguchi et al., 2009). It is composed of 10 bidirectional transcribed open reading frames arranged in two putative operons (*mcy* A-C and *mcy* D-J) by a promotor region (Tillett et al., 2000; Zurawell et al., 2005). The *mcy* gene cluster cannot only be used as gene marker to quantify toxic genotypes, such as *mcy* A, *mcy* B, *mcy* D, *mcy* E and *mcy* J (Zuo et al., 2018), but can also provide a new tool for the investigation of microcystin variation, evolution and function (Mikalsen et al., 2003). For the reason that the *mcy* gene cluster does not well tolerate mutations with respect to toxin biosynthesis (Pearson et al., 2006), microcystin-production cyanobacteria often produce several isoforms of microcystin (Mikalsen et al., 2003; Pearson et al., 2006). In addition, variability is also probably influenced by environmental conditions (Davis, 2009). Among these variants, MC-LR (with leucine and arginine amino acids), a potent inhibitor of protein phosphatase and inducer of cytoskeleton alterations, is one of the most common and toxic variants of MCs (Oudra et al., 2002; Chen et al., 2016).

In natural freshwater suffering harmful cyanobacteria blooms, it is important to investigate content and variation of MCs for their widespread occurrence, acute toxicity and tumor-promoting property. Moreover, identifying the factors influencing the MCs production ability of toxic cyanobacteria is profitable for the control of MCs pollution in natural freshwater. As an important part of natural freshwater, aquaculture water experiences much less hydrologic changes and much more impacts of aquaculture activities compared to other freshwater bodies. Cyanobacteria blooms occur more frequently in aquaculture water (Hu et al., 2018). MCs produced in cyanobacteria blooms can be ingested and accumulated in the aquacultured animals, and eventually endanger human health via the food chain (Codd et al., 2005). In this paper, a field study in MC-LR content and the abundance of *mcy* genes was performed in five man-made freshwater fishponds suffering cyanobacteria blooms. Moreover, we respectively analyzed the relationship between the MC-LR production ability of toxic cyanobacteria and MC-LR content, TN content, TP content, SNP of the metagenomic *mcy* genes.

Materials and Methods

Study sites and sampling

5 freshwater fish ponds (warm water) with a surface area of 3500-7000 m^2 and mean water depth of 1.5 m were located in Xiqing district (XQ, site N38°57′20.74″, E117°11′24.43″), Ninghe district (NH, site N39°26′30.81″, E117°29′46.52″), Dagang district (DG, site N38°48′14.81″, E117°28′59.48″), Jizhou district (JX, site N39°50′19.24″, E117°20′46.21″) and Beichen district (BC, site N39°17′0.24″, E117°24′0.38″) in Tianjin, China, respectively (*Fig. 1*).

Sampling was conducted during July 2018 and performed according the method of Bi et al. (2019). Briefly, to collecting mixed water sample from the surface to the bottom, a specially designed cylindrical organic glass sampler was used. Water samples were collected three times per sampling site. The mixed water sample using in the chemical analysis (2 L) was stored at 4 °C and transferred to the laboratory condition within 2 h. The mixed water sample using in DNA extraction (approximately 500 mL) was filtered through 0.22 μm millipore polycarbonate membrane, and then the membrane was immediately stored at -20 °C. The mixed water sample (1 L) was filtered through 0.45

http://www.aloki.hu ● ISSN 1589 1623 (Print) ● ISSN1785 0037 (Online)
DOI: http://dx.doi.org/10.15666/aeer/1901_237248

μm and 0.22 μm cellulose acetate membranes sequentially, in which the filtered water sample and the seston retained on the membrane were using in MC-LR content analysis, respectively.

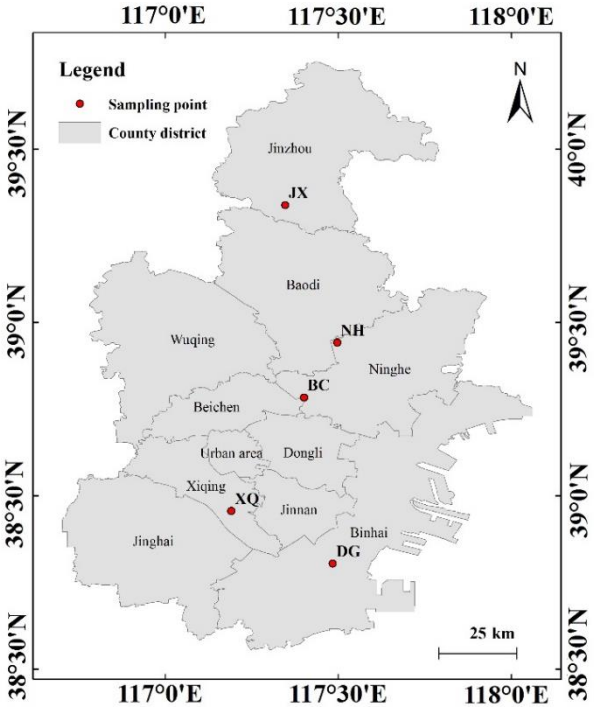

Figure 1. *Map of sampling points in Tianjin, China*

Total nitrogen and total phosphorus contents analysis

TN and TP contents of the mixed water sample were detected via the alkaline potassium persulfate digestion method (China National Standard GB11894-89) and ammonium molybdate spectrophotometric method with or without potassium persulfate digestion (China National Standard GB11893-89), respectively.

MCs extraction and MC-LR determination

Supelco SPE C_{18} cartridge was preconditioned with 10 mL methanol and 10 mL Milli-Q (Millipore, UK) water and used to enrich MCs of the filtered water sample. MCs in cartridge was eluted with 20% aqueous methanol, and the elution was evaporated on a rotary evaporator to 0.5-1 mL and dried with nitrogen gas. The residue was redissolved in 1 mL of 50% methanol and filtered through a 0.22 μm nylon membrane. After 4 freeze-thaw cycles, the seston was lyophilized at -40 °C under 8-11 mTorr and extracted three times with 20 mL of 90% aqueous methanol. The extraction liquid was centrifugated at 12000 g for 10 min. The mixed supernatant was evaporated to 0.5-1 mL, and then was added 6 mL Milli-Q to extract MCs using a SPE C_{18} cartridge as described above.

Above concentrated MCs samples were analyzed via a HPLC system (SPD-M20A, Shimadzu, Japan) equipped with a Shim-Pack VP-ODS column (250 mm◊4.6 mm) and a DAD detector, using 60% aqueous methanol with 0.05% trifluoroacetic acid at a flow rate of 1 mL/min. MC-LR were identified by the retention time, and quantification was calculated based on the standard curve of certified MC-LR standards (Sigma, USA). The

http://www.aloki.hu ● ISSN 1589 1623 (Print) ● ISSN1785 0037 (Online)
DOI: http://dx.doi.org/10.15666/aeer/1901_237248

sum of MC-LR contents dissolved in filtrate and retained in seston represents MC-LR content produced by the metagenomic *mcy* gene cluster of the water sample.

DNA extraction and shotgun library preparation and sequencing

Total genome DNA from samples was extracted using DNeasy PowerSoil (QIAGEN, U.S.) according to the manufacturer's instructions and eluted in 50 μL of elution buffer. The harvested DNA was detected by the agarose gel electrophoresis and quantified by Qubit (Thermo Fisher, U.S.).

The metagenome DNA libraries were constructed with 1 μg of DNA genomes, according to the Illumina TruSeq DNA Sample Prep v2 Guide (Illunima), with an average of 350 bp insert size. All qualified libraries were loaded to Hiseq X ten to perform pair-end sequencing. More than 10G raw reads were obtained for each sample.

Bioinformatic analysis

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 base (Phred score≥20) were removed; (3) 3'-ends of reads were trimmed to the first high-quality base. Compared to reference of *M. aeruginosa* NIES-843, high-quality reads obtained through mapping metagenomic reads using BWA (https://nchc.dl.sourceforge.net/project/bio-bwa/bwa-0.7.15.tar.bz2) were assembled with multiple Kmer parameters using SOAPdenovo (version 2.04) (http://soap.genomics.org.cn/) to obtain the optimal assembled results. Local cavity filling and base correction were performed on the assembled results with GapCloser (version 1.12). The final assembled results are shown in *Table 1*. Assembled results were mapped to the *mcy* gene cluster of reference gene of *M. aeruginosa* NIES-843 using BWA, and bam files were obtained using samtools. The abundance of *mcy* gene A-J of samples were obtained using humann2 (http://huttenhower.sph.harvard.edu/humann2), and SNP identification of metagenomic *mcy* gene was using VarScan (http://varscan.sourceforge.net/). All sequences included in this paper were submitted to GenBank (Accession Number: SRP255704).

Table 1. *The assembly results were statistically analyzed*

Sampling points	Jizhou (JX)	Ninghe (NH)	Xiqing (XQ)	Beichen (BC)	Dagang (DG)
No. of all contigs	35	1070	653	441	215
Bases in all contigs (bp)	60245	268446	226925	192314	112172
Average length (bp)	1721.3	250.9	347.5	436.1	521.7
No. of large contigs (>1000 bp)	20	4	13	19	19
Bases in large contigs	53295	5120	16395	24365	36064
Largest length (bp)	6500	1904	1598	1579	4757
Contig N50 (bp) (>1000 bp)	3302	1136	1275	1348	1915
Contig N90 (bp) (>1000bp)	1342	1026	1032	1049	1138
G+C content (%)	38.95	38.90	38.99	39.10	38.88

Data analysis

The correlation analyses were conducted using SPSS (version 17.0).

http://www.aloki.hu ● ISSN 1589 1623 (Print) ● ISSN1785 0037 (Online)
DOI: http://dx.doi.org/10.15666/aeer/1901_237248

Results

Relationship between the metagenomic mcy gene abundance and MC-LR content, TN and TP contents

MC-LR content produced by toxic cyanobacteria and the abundance of metagenomic *mcy* genes in five fishponds were presented in *Table 2*. MC-LR content and the abundance of 10 metagenomic *mcy* genes varied with sample sites. Except for the metagenomic *mcy* H and I at sampling point NH, MC-LR content and the abundance of 10 metagenomic *mcy* genes increased in the order of JX, NH, XQ, BC, and DG. As shown in *Table 3*, there was positive correlation between MC-LR content and all metagenomic *mcy* gene abundance. MC-LR content had significant correlation with the metagenomic *mcy* A-C abundance ($P<0.05$) and extremely significant correlation with the metagenomic *mcy* E-J abundacne ($P<0.01$). There was indistinctive correlation between the metagenomic *mcy* A-J abundance and contents of TN and TP ($P>0.05$).

Table 2. *MC-LR content and the abundance of metagenomic mcy gene in five sampling sites*

Sampling sites	MC-LR content (μg/L)	Gene abundance (Copies/L)									
		***mcy* C**	***mcy* B**	***mcy* A**	***mcy* D**	***mcy* E**	***mcy* F**	***mcy* G**	***mcy* H**	***mcy* I**	***mcy* J**
JX	12.56	189.44	236.6	231.28	501.2	343.05	29.46	321.63	114.15	45.39	32.76
NH	13.86	11135.14	18915.5	25456.45	10441.56	5743.14	526.67	9302.17	0	0	603.36
XQ	17.82	12458.23	21608.17	25892.17	40107.65	35942.4	2443.72	24634.97	6735.61	3614.12	3383.46
BC	20.55	16166.58	28035.62	33157.23	51039.3	44564.36	2708.76	31237.86	6893.65	4705.23	3858.43
DG	23.86	28337.19	48898.12	59991.29	88380.13	80683.54	6080.56	54817.68	14168.78	7906.41	7274.32

Table 3. *Correlation between the abundance of metagenomic mcy A-J and MC-LR, TN and TP contents*

	***mcy* C**		***mcy* B**		***mcy* A**		***mcy* D**		***mcy* E**		***mcy* F**		***mcy* G**		***mcy* H**		***mcy* I**		***mcy* J**	
	r	**D.W**	**r**	**D.W**	**r**	**D.W**	**r**	**D.W**	**r**	**D.W**	**r**	**D.W**	**r**	**D.W**	**r**	**D.W**	**r**	**D.W**	**r**	**D.W**
MC-LR	0.93*	2.42	0.94*	2.43	0.91*	2.34	0.99**	2.68	0.99**	2.75	0.96**	2.33	0.99**	2.51	0.97**	2.96	0.99**	3.05	0.98**	2.85
TN	0.13	0.81	0.14	0.80	0.10	0.87	0.26	0.59	0.26	0.61	0.18	0.73	0.23	0.61	0.29	0.73	0.30	0.58	0.27	0.65
TP	0.81	1.79	0.81	1.76	0.81	1.93	0.80	1.33	0.81	1.34	0.87	1.37	0.81	1.34	0.83	1.38	0.77	1.38	0.82	1.31

* Correlation is significant at $P<0.05$, ** Correlation is significant at $P<0.01$, 'r' is the correlation coefficient and 'D.W.' is the test value of Durbin-Watson

Relationship between the MC-LR production ability of all metagenomic mcy genes and MC-LR content, TN and TP contents

As shown in *Table 4*, the natural log of the ratio of MC-LR content to the metagenomic *mcy* gene abundance (Ln MC-LR/*mcy*) reflecting the MC-LR production ability of the metagenomic *mcy* genes was calculated. The MC-LR production ability of all metagenomic *mcy* genes at sampling point JX was the highest in 5 sampling points. As shown in *Table 5*, the MC-LR production ability of the metagenomic *mcy* gene A-J was in indistinctive correlation with the abundance of metagenomic *mcy* gene A-J and MC-LR content and TN and TP contents ($P>0.05$).

http://www.aloki.hu ● ISSN 1589 1623 (Print) ● ISSN1785 0037 (Online)
DOI: http://dx.doi.org/10.15666/aeer/1901_237248

Table 4. *The natural log of the ratio of MC-LC content to the metagenomic mcy gene abundance (Ln MC-LR/mcy)*

Sampling points	Ln MC-LR/*mcy*									
	***mcy* C**	***mcy* B**	***mcy* A**	***mcy* D**	***mcy* E**	***mcy* F**	***mcy* G**	***mcy* H**	***mcy* I**	***mcy* J**
JX	-2.71	-2.94	-2.91	-3.69	-3.31	-0.85	-3.24	-2.21	-1.28	-0.96
NH	-6.89	-7.22	-7.52	-6.62	-6.03	-3.64	-6.51			-3.77
XQ	-6.55	-7.1	-7.28	-7.72	-7.61	-4.92	-7.23	-5.93	-5.31	-5.25
BC	-6.67	-7.22	-7.39	-7.82	-7.68	-4.88	-7.33	-5.82	-5.43	-5.24
DG	-7.08	-7.63	-7.83	-8.22	-8.13	-5.54	-7.74	-6.39	-5.8	-5.72

Table 5. *The correlation between Ln MC-LR/mcy and the metagenomic mcy gene abundance and MC-LR, TN and TP contents*

	Ln MC-LR/*mcy* C		**Ln MC-LR/*mcy* B**		**Ln MC-LR/*mcy* A**		**Ln MC-LR/*mcy* D**		**Ln MC-LR/*mcy* E**	
	r	**D.W**	**r**	**D.W**	**r**	**D.W**	**r**	**D.W**	**r**	**D.W**
Gene abundance	-0.79	1.78	-0.80	1.71	-0.81	1.67	-0.81	1.67	-0.82	1.57
MC-LR content	-0.64	1.26	-0.67	1.30	-0.66	1.27	-0.83	1.39	-0.86	1.34
TN	-0.38	1.91	-0.41	1.83	-0.40	1.87	-0.54	1.25	-0.57	1.06
TP	-0.43	1.76	-0.45	1.67	-0.44	1.71	-0.52	1.23	-0.54	1.14
	Ln MC-LR/*mcy* F		**Ln MC-LR/*mcy* G**		**Ln MC-LR/*mcy* H**		**Ln MC-LR/*mcy* I**		**Ln MC-LR/*mcy* J**	
	r	**D.W**	**r**	**D.W**	**r**	**D.W**	**r**	**D.W**	**r**	**D.W**
Gene abundance	-0.79	1.56	-0.79	1.74	-0.87	2.01	-0.88	2.04	-0.82	1.62
MC-LR content	-0.85	1.40	-0.79	1.40	-0.90	1.73	-0.90	1.65	-0.85	1.34
	-0.53	1.15	-0.50	1.43	-0.56	0.65	-0.56	0.61	-0.56	1.13
	-0.57	1.20	-0.51	1.35	-0.56	1.62	-0.52	1.59	-0.54	1.16

'r' is the correlation coefficient and 'D.W.' is the test value of Durbin-Watson

Relationship between the MC-LR production ability of the metagenomic mcy genes and SNPs

As shown in *Table 6* and *Figure 2*, at some single nucleotide sites in the metagenomic *mcy* gene A-J, SNPs presented significant (P<0.05) /extremely significant (P<0.01) correlation with the MC-LR production ability of the metagenomic *mcy* genes where 18 SNPs and 388 SNPs were strongly positively correlated and strongly negatively correlated with the MC-LR production ability of the metagenomic *mcy* genes, respectively. No significantly positive correlation between SNP and the MC-LR production ability in the metagenomic *mcy* B and C was observed. In the metagenomic *mcy* C, the number of SNPs significantly negatively correlated with the MC-LR production ability was 5, and the percentage of SNP significantly correlated with the MC-LR production ability/ the length of the metagenomic *mcy* gene was the lowest in NPRSs. Meanwhile, the percentage of SNPs significantly correlated with the MC-LR production ability for amino acid change / SNPs in the metagenomic *mcy* gene was the lowest in all metagenomic *mcy* genes. SNP for amino acid change in SNPs significantly correlated with the MC-LR production ability was not observed in *mcy* I and *mcy* J. The amino acid sites changed by SNPs significantly correlated with the MC-LR production ability in the metagenomic *mcy* A-J were shown in *Table 7*, respectively.

http://www.aloki.hu ● ISSN 1589 1623 (Print) ● ISSN1785 0037 (Online)
DOI: http://dx.doi.org/10.15666/aeer/1901_237248

Table 6. *Sites in the metagenomic mcy gene A-J that SNP was significantly/ extremely significantly correlated with the MC-LR production ability*

Reference (NIES-843) Gene (length in base pair)	Correlation	The mutant single nucleotide site								
mcy C (3876)	Positive									
	Negative	S1213*	***S2204****	***S3379****	***S3566****	S3604*				
mcy B (6381)	Positive									
	Negative	***S4309****	S4938*	S4941*	***S4946****	S5163**	S5184**	S5226**	***S5240****	***S5246****
		S5250*	S5508**	S5664**	S5748*	S7482**	S7485**	***S7488****	S7650*	S7668**
		S7713*	S7716*	***S8062****	***S8320****	S8547*	***S9708****	S9709*	***S9911****	
mcy A (8364)	Positive	***S18576****								
	Negative	S10317**	***S10553****	***S11119****	S11127**	S11319*	S11481*	***S11686****	S12111*	S12141*
		S12636*	S12711*	S13890**	S14274**	S14280**	***S14459****	***S15559****	***S15575****	***S15676****
		S15834*	S16034*	S16293*	S16305**	S16404*	S16860*	S17247**	S17274*	***S17801****
		S18019*	***S18068****	S18156*						
mcy D (11706)	Positive	S27665**	***S28915****	S30725*	***S30934****	S30938*	S30941*	***S31071****		
	Negative	***S19655****	S19673**	S19733**	***S19749****	S19850**	S19928**	S20096**	S20108**	S20177**
		S20183**	***S20415****	S20438**	S20474**	S20504**	S20507**	***S20513****	S20561**	***S20643****
		S20654**	S20666**	S20717**	S20783*	S21086*	S21101**	***S21160****	***S21161****	S21167*
		S21185*	***S21261****	***S21270****	***S21285****	***S21288****	***S21294****	***S21307****	***S21322****	***S21342****
		S21348*	***S21353****	***S22011****	***S22491****	***S22518****	***S22520****	S22685*	S22697**	S22712**
		S22731*	S22766*	S22772**	***S22773****	S22787**	***S22882****	S22886*	S22937*	***S22991****
		S23018**	***S23049****	***S23101****	S23123*	S23126*	S23144*	***S23145****	***S23146****	S23153**
		S23191*	***S23328****	***S23334****	S23414**	S23438*	S23441*	***S23443****	***S23465****	S23528*
		S23571*	***S23593****	***S23598****	S24137*	***S24141****	S24144*	S24170**	***S24172****	***S24173****
		S24187*	***S24279****	***S24398****	S24470*	***S24633****	S24722*	S24794*	***S24769****	S24849*
		S24906*	***S24927****	***S24972****	***S24974****	S24978*	***S25005****	S25010*	***S25014****	***S25071****
		S26042**	S26312**	S26384**	S26465**	S26630**	S26858**	S27356*	***S27396****	S27929**
		S27965*	S28034*	S28058*	S28106**	S28106**	***S28171****	***S28209****	S28292*	S28295*
		S28301*	***S28500****	***S28501****	S28625**	S28688**	***S28777****	S29153**	S29237*	***S29238****
		S29339*	S29703*	***S29819****	***S29821****	S29852*	***S29950****	***S30052****	S30119*	S30221**
		S30335**	***S30401****	S30491*	***S30709****					
mcy E (10464)	Positive	S38712*								
	Negative	S31617**	S31692*	S31743**	S31854**	S31893**	S32013**	S32130**	S32502**	***S32515****
		S32568*	***S32628****	S32748**	S32943**	***S33592****	S33645**	S33903**	S34608**	S34770*
		S34878*	S34932**	S35223**	S35232**	***S35234****	S35253**	S35316*	S35328**	***S35594****
		S35595*	***S35596****	***S35602****	***S35604****	***S35785****	S35799*	***S35813****	***S35814****	S35889**
		S36009**	S36063**	S36078**	S36438**	***S36439****	S36454**	S36462*	***S36484****	S36924**
		S36945**	S37050**	S37089**	S37356*	S37386**	***S37457****	***S37531****	***S37570****	***S37590****
		S37593**	***S37597****	S37672**	S37686**	***S37717****	S37734**	***S37950****	S37953*	S38037*
		S38139**	S38190**	S38280**	***S38467****	***S38468****	S38631*	S38715*	S38838*	S39303*
		S39555*	S40786*	S40839*	***S40919****	***S40961****	S41010*	S41112*	***S41144****	S41169*
		S41319*	S41328**	***S41407****	***S41473****	S41535*	S41553*	***S41596****	S41625**	

http://www.aloki.hu ● ISSN 1589 1623 (Print) ● ISSN1785 0037 (Online)
DOI: http://dx.doi.org/10.15666/aeer/1901_237248

mcy F (756)	Positive	***S42113****								
	Negative	S42196*	S42244*	***S42415****	S42616*	***S42618****	***S42621****	S42637*	***S42645****	
mcy G (7896)	Positive	S44772**	***S47707****							
	Negative	S43053*	S43191*	***S43223****	***S43316****	S43350*	S43416*	S43464*	***S43519****	S43536*
		S43563*	S43611*	S43647*	***S43664****	S44730*	***S44732****	***S44777****	S44904*	S45048*
		S45066*	S45099**	***S45113****	S45171*	S45321*	S45393*	S45408*	S45415*	S45459*
		S45480*	S45495*	***S45559****	S45666*	S45669*	S46089**	S46197**	S46203**	***S46403****
		S46452**	***S46530****	S46560**	S46686*	S46695*	S46716*	S46734*	***S46769****	***S46857****
		S46866*	S46869*	S46872*	***S46934****	S46962*	S46977*	S46992*	S47001*	S47028*
		S47631*	***S47656****	S47706*	***S47719****	S47725*	***S47731****	***S47732****	S47742*	***S47775****
		S47950*	***S47953****	S48312*	***S48886****	S49410**	S49833*	***S49895****	***S50017****	***S50018****
		S50019*	S50022*	S50070*	S50106*	S50139*	S50175*	S50187*	***S50322****	***S50325****
		S50331*	S50340**	***S50512****						
mcy H (1758)	Positive	***S51149****	***S51167****	S51175**	S51181**	S51304**	S51353**			
	Negative	S51271**	***S51275****	***S51371****	***S51393****	***S51611****	***S51828****	***S51831****	***S51875****	***S51881****
		S52294*								
mcy I (1014)	Positive									
	Negative	S53559**								
mcy J (936)	Positive									
	Negative	S54355**								

* Correlation is significant at $P<0.05$, ** Correlation is significant at $P<0.01$, 'S' means the site of SNP in the metagenomic *mcy* gene cluster and the site of SNP in italics and bold means SNP for amino acid change

Table 7. *The site of amino acid changed in the metagenomic mcy A-J by SNP significantly correlated with the MC-LR production ability*

***Mcy* gene**	**The mutant site of amino acid**									
mcy C	104	166	558							
mcy B	115	181	645	731	922	1668	1670	1672	1770	
mcy A	19	189	205	278	986	1020	1025	1392	2316	***2505***
mcy D	39	71	293	325	369	541	549	575	578	583
	586	590	595	602	604	605	825	985	994	1065
	1115	1151	1171	1188	1203	1218	1264	1266	1302	1309
	1351	1353	1535	1545	1550	1581	1620	1699	1744	1790
	1823	1826	1845	***2620***	2878	2891	2988	3080	3126	3234
	3428	3471	3605	3621	***3724***	3795	***3845***			
mcy E	369	406	728	1156	1275	1395	1396	1398	1459	1468
	1692	2016	2041	2054	2060	2063	2103	2180	2353	3184
	3245	3333	3355	3396						
mcy F	***70***	170	238	239	247					
mcy G	144	175	243	291	647	662	774	923	1204	1246
	1355	1381	1623	1640	1644	1647	1661	1720	1721	2032
	2409	2510	2511	2513	2574					
mcy H	***127***	***133***	169	201	208	281	353	354	369	

The amino sites in italics and bold were changed by SNPs positively correlated with the MC-LR production ability. The amino sites in normal font were changed by SNPs negatively correlated with the MC-LR production ability

http://www.aloki.hu ● ISSN 1589 1623 (Print) ● ISSN1785 0037 (Online)
DOI: http://dx.doi.org/10.15666/aeer/1901_237248

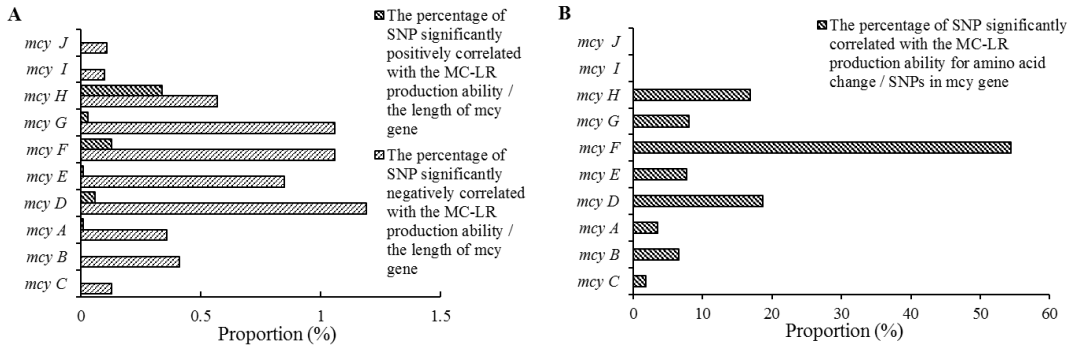

Figure 2. *The proportion of SNP significantly correlated with the MC-LR production ability of mcy genes*

Discussion

Toxic cyanobacteria likely carried only one copy of the *mcy* gene cluster genetically structured similarly as the housekeeping genes and therefore the *mcy* gene cluster abundance was used to predict potential pollution of MCs in natural freshwater (Kyoung-Hee et al., 2012; Zuo et al., 2018), and most methods of detecting and identifying MCs producers were based on PCR using primers designed to recognize the *mcy* gene cluster (Rantala et al., 2006). In the present study, there was a strong positive correlation between MC-LR content and the abundance of metagenomic *mcy* gene A-J, while an indistinctive correlation between the MC-LR production ability of metagenomic *mcy* gene and abundance of metagenomic *mcy* gene and the MC-LR content was observed. It was suggested that we should take a closer look at the indicators predicting potential pollution of MCs in natural freshwater. Zuo et al. (2018) suggested that the ratio of toxic *Microcystis* determined by *mcy* B abundance was higher in both laboratory and field samples, which might result from sequence characteristics of *mcy* B amplicon. *mcy* B might be amplified more easily in a competitive PCR system. This characteristic of *mcy* B was not observed in the present study. *mcy* H are putatively involved in location and stabilizing the megasynthase (Tillett et al., 2000; Liu et al., 2019). *mcy* I may play a role in dehydration and stabilization of the microcystin synthase complex (Tillett et al., 2000; Liu et al., 2019). Notably, both *mcy* H and I abundance at site NH were 0, whereas MC-LR content was 13.86 μg/L. A similar result was also reported by Pearson et al. (2006), who found *Planktothrix agardhii* CYA126 without *mcy* I was capable of production MCs. The reason for this phenomenon might be that both *mcy* H and *mcy* I was not critical for the biosynthesis of MC-LR.

Cyanobacteria blooms in freshwater are broadly associated with eutrophic and poorly flushed water. As nutrients in freshwater bodies, especially phosphorus, become enriched, there is often a shift in the phytoplankton community towards dominance by cyanobacteria (Davis, 2009). The regulation factors as a whole governing MCs synthesis remain unknown (Kuniyoshi et al., 2013). Previous researches suggested the expression of the *mcy* gene cluster could be regulated by environmental factors and toxic cyanobacterial strains, which appear to have higher N and P requirements than nontoxic strains, possibly due to the extra energy and materials required for toxin synthesis (Zurawell et al., 2005; Davis, 2009). According to previous research, the correlation between MCs content and various P level was to be positive or negative (Davis, 2009), and it was hypothesized that the effects of nitrogen on toxin-production ability of

http://www.aloki.hu ● ISSN 1589 1623 (Print) ● ISSN1785 0037 (Online)
DOI: http://dx.doi.org/10.15666/aeer/1901_237248

cyanobacteria depended on the limitation of phosphorus (Zurawell et al., 2005; Kuniyoshi et al., 2013). Under the P-limited condition, a negative correlation was shown between orthophosphate and TP, and MCs (Kuniyoshi et al., 2013). Content of total MCs, in particular the more toxic MC-LR variant, increased with increasing P limitation (Zurawell et al., 2005). Sevilla et al. (2010) suggested that both the transcription level of *mcy* D and MC-LR per cell had been shown to be independent of nitrate availability in sufficient P conditions. In the present study, there was higher level of both TP content (>0.10 mg/L) and TN content (>0.85 mg/L) at all 5 sampling sites, and both the abundance and the MC-LR production ability of all metagenomic genes were in indistinctive correlation with TP content, coinciding with the result of Sevilla et al. (2010). Davis (2009) suggested that toxic strains of *Microcystis* were able to outgrow non-toxic strains at high N levels. However, we found an indistinctive correlation between TN content and both the abundance and the MC-LR production ability of all metagenomic *mcy* genes.

In the *mcy* gene cluster, NRPSs including *mcy* A-C, E catalyzed the formation of peptides, and PKS including *mcy* G, E and D were involved in the formation of Adda, and *mcy* F, H-J encoded modifying genes involved in epimerization, localization, dehydration and O-methylation (Tillett et al., 2000; Zurawell et al., 2005). Attributed to the relaxed substrate specificity of the adenylation domain and genetic variation in the *mcy* gene cluster, toxic *Microcystis* strains often produce several isoforms of the cyclic hepatotoxin microcystin (Mikalsen et al., 2003). The main characteristic of MC-LR was L-leucine at the variable amino acid position X2 and especial L-arginine at the variable amino acid position Z4, in which the *mcy* B1 module and the *mcy* C module were respectively involved in recognition of specific substrate of these amino acid position (Tanabe et al., 2009). In present study, no SNP significantly positively correlated with the MC-LR production ability was found in *mcy* B and *mcy* C. It was predicted that the genotype of *mcy* C and *mcy* B in *M. aeruginosa* NIES-843 was the most favorable for MC-LR production in present study. Some SNPs significantly positively correlated with MC-LR production ability of metagenomic *mcy* genes were observed in *mcy* A and *mcy* D-H, suggesting these sites just were critical to MCs production according to the role of the *mcy* genes including these sites in MCs synthesis. In the present study, *mcy* C had the lowest proportion of SNPs significantly correlated with MC-LR production ability in NRPSs and the lowest proportion of SNPs significantly correlated with MC-LR production ability for amino acid change in all metagenomic *mcy* genes. It was indicated that *mcy* C was more critical to the production of MC-LR and there might be some mechanism to avoid substrate specific changes due to genetic mutation. *mcy* genes belonging to PKS had higher mutation significantly negatively correlated with the MC-LR production ability of the metagenomic *mcy* genes, predicting that *mcy* D, E and G played important roles in MCs synthesis according to the role of PKS in MCs synthesis.

Conclusion

In the present study, MC-LR content was positively correlated with the abundance of metagenomic *mcy* gene A-J. The MC-LR production ability of all metagenomic *mcy* genes was in indistinctive correlation with the metagenomic *mcy* gene abundance, and MC-LR content, and TN and TP content in sufficient N and P conditions. Both *mcy* C and *mcy* B of *M. aeruginosa* NIES-843 had the strongest MC-LR production ability in the metagenomic *mcy* C and B of five sampling sites. *mcy* C was the most critical to the production of MC-LR in all *mcy* genes and there might be some mechanism to avoid

http://www.aloki.hu ● ISSN 1589 1623 (Print) ● ISSN1785 0037 (Online)
DOI: http://dx.doi.org/10.15666/aeer/1901_237248

substrate specific changes due to genetic mutation. It was necessary to take a closer look at the indicators predicting potential pollution of MCs in natural freshwater. Using genome-editing techniques to further identify critical SNPs correlated with the MC-LR production ability would offer novel insight into the correlation among SNP genotypes composition and toxin characteristic of microcystin-production cyanobacteria in aquaculture ponds.

Acknowledgments. This work was supported by the National Natural Science Foundation of China (grant no. 31772857), Natural Science Foundation Grant of Tianjin (grant nos. 18JCYBJC95900 and 20JCZDJC23600, 17JCYBJC29500), and Key Research and Development Projects of Tianjin (grant no. 19YFZCSN00070).

REFERENCES

[1] Bi, X. D., Dai, W., Wang, X. Y., Dong, S. J., Zhang, S. L., Zhang, D. J., Wu, M. (2019): Microcystins distribution, bioaccumulation, and *Microcystis* genotype succession in a fish culture pond. – Science of the Total Environment 688: 380-388.

[2] Carmichael, W. W. (2001): Health effects of toxin-producing cyanobacteria: "the Cyano-HABs". – Human Ecological Risk Assessment 7: 1393-1407.

[3] Chen, L., Chen, J., Zhang, X. Z., Xie, P. (2016): A review of reproductive toxicity of microcystins. – Joural of Hazardous Materials 301: 381-399.

[4] Codd, G. A., Morrison, L. F., Metcalf, J. (2005): Cyanobacterial toxins: risk management for health protection. – Toxicology Applied Pharmacology 203(3): 264-272.

[5] Davis, T. (2009): Effects of nutrients, temperature, and zooplankton grazing on toxic and non-toxic strains of the harmful cyanobacterium *Microcystis* spp. – the Degree of Doctor of Philosophy, Stony Brook University.

[6] Falconer, I. R. (1999): An overview of problems caused by toxic blue-green algae (cyanobacteria) in drinking and recreational water. – Environmental Toxicology 14: 5-12.

[7] Hu, X. B., Zhang, R. F., Ye, J. Y., Wu, X., Zhang, Y. X., Wu, C. L. (2018): Monitoring and research of microcystins and environmental factors in a typical artificial freshwater aquaculture pond. – Environmental Science and Pollution Research 25: 5921-5933.

[8] Kuniyoshi, T. M., Sevilla, E., Bes, M. T., Fillat, M. F., Peleato, M. L. (2013): Phosphate deficiency (N/P 40:1) induces *mcyD* transcription and microcystin synthesis in *Microcystis aeruginosa* PCC7806. – Plant Physiology and Biochemistry 65: 120-124.

[9] Liu, T. Z., Mazmouz, R., Pearson, L. A., Neilan, B. A. (2019): Mutagenesis of the microcystin tailoring and transport proteins in a heterologous cyanotoxin expression system. – Synthetic Biology 8: 1187-1194.

[10] Mikalsen, B., Boison, G., Skulberg, O. M., Fastner, J., Davies, W., Gabrielsen, T. M., Rudi, K., Jakobsen, K. S. (2003): Natural variation in the microcystin synthetase operon *mcyABC* and impact on microcystin production in *Microcystis* strains. – Journal of Bacteriology 185(9): 2774-2785.

[11] Ngwa, F. F., Madramootoo, C. A., Jabaji, S. (2014): Comparison of cyanobacterial microcystin synthetase (*mcy*) E gene transcript levels, *mcy* E gene copies, and biomass as indicators of microcystin risk under laboratory and field conditions. – MicrobiologyOpen 3(4): 411-425.

[12] Noguchi, T., Shinohara, A., Nishizawa, A., Asayama, M., Nakano, T., Hasegawa, M. (2009): Genetic analysis of the microcystin biosynthesis gene cluster in microcystis strains from four bodies of eutrophic water in japan. – Journal of General & Applied Microbiology 55(2): 111-123.

http://www.aloki.hu ● ISSN 1589 1623 (Print) ● ISSN1785 0037 (Online)
DOI: http://dx.doi.org/10.15666/aeer/1901_237248

[13] Oh, K-H., Jeong, D-H., Cho, Y-C. (2012): Quantification of toxigenic *Microcystis* spp. In freshwaters by quantitative real-time PCR based on the microcystin synthetase A gene. – Joural of Microbiology 51(1): 18-24.
[14] Oudra, B., Loudiki, M., Sbiyyaa, B., Sabour, B., Martins, R., Amorim, A., Vasconcelos, V. (2002): Detection and variation of microcystin contents of Microcystis blooms in eutrophic Lalla Takerkoust Lake, Morocco. – Lakes & Reservoirs: Research and Management 7: 35-44.
[15] Pearson, L. A., Barrow, K. D., Neilan, B. A. (2006): Characterization of the 2-hydroxy-acid dehydrogenase *mcy* I, encoded within the microcystin biosynthesis gene cluster of *Microcystis aeruginosa* PCC7806. – The Journal of Biological Chemistry 282(7): 4681-4692.
[16] Rantala, A., Rajaniemi-Wacklin, P., Lyra, C., Lepistö, L., Rintala, J., Mankiewicz-Boczek, J., Sivonen, K. (2006): Detection of microcystin-producing cyanobacteria in Finnish lakes with genus-specific microcystin synthetase gene E (*mcy E*) PCR and associations with environmental factors. – Applied and Environmental Microbiology 72(9): 6101-6110.
[17] Rastogi, R. P., Sinha, R. P., Incharoensakdi, A. (2014): The cyanotoxin-microcystins: current overview. – Reviews in Environmental Science and Biotechnology 13: 215-249.
[18] Sevilla, E., Martin-Luna, B., Vela, L., Bes, M. T., Peleato, M. L., Fillat, M. F. (2010): Microcystin-LR synthesis as response to nitrogen: transcriptional analysis of the *mcy*D gene in *Microcystis aeruginosa* PCC7806. – Ecotoxicology 19: 1167-1173.
[19] Tanabe, Y., Sano, T., Kasai, F., Watanabe, M. M. (2009): Recombination, cryptic clades and neutral molecular divergence of the microcystin synthetase (*mcy*) genes of toxic cyanobacterium *Microcystis aeruginosa*. – BMC Evolutionary Biology 9: 115.
[20] Tillett, D., Dittmann, E., Erhard, M., von Döhren, H., Börner, T., Neilan, B. A. (2000): Structural organization of microcystin biosynthesis in *Microcystis aeruginosa* PCC7806: an integrated peptide-polyketide synthetase system. – Chemistry & Biology 7: 753-764.
[21] Zuo, J., Chen, L. T., Shan, K., Hu, L. L., Song, L. R., Gan, N. Q. (2018): Assessment of different *mcy* genes for detecting the toxic to non-toxic *Microcystis* ratio in the field by multiplex q PCR. – Journal of Oceanology and Limnology 36(4): 1132-1144.
[22] Zurawell, R. W., Chen, H. R., Burke, J. M., Prepas, E. E. (2005): Hepatotoxic cyanobacteria: a review of the biological importance of microcystins in freshwater environments. – Journal of Toxicology and Environmental Health, Part B 8: 1-37.

http://www.aloki.hu ● ISSN 1589 1623 (Print) ● ISSN1785 0037 (Online)
DOI: http://dx.doi.org/10.15666/aeer/1901_237248