US 20070059699A1

(19) **United States**

(12) **Patent Application Publication** (10) **Pub. No.: US 2007/0059699 A1**

Sivonen et al. (43) **Pub. Date: Mar. 15, 2007**

(54) **METHODS FOR DETECTING TOXIC AND NON-TOXIC CYANOBACTERIA**

(76) Inventors: **Kaarina Sivonen**, Helsinki (FI); **Anne Rantala**, Helsinki (FI); **Leo Rouhianinen**, Helsinki (FI); **David Fewer**, Helsinki (FI); **Pirjo Rajaniemi**, Helsinki (FI); **Annick Wilmotte**, Tiliff (BE); **Christophe Boutte**, Le Charnbon Feugerolles (FR); **Stana Grubisic**, Liege (BE); **Pierre Balthasart**, Liege (BE); **Gianluca De Bellis**, Segrate Milano (IT); **Ermanno Rizzi**, Segrate Milano (IT); **Andrea Frosini**, Segrate Milano (IT); **Bianca Castiglioni**, Segrate Milano (IT); **Stefano Ventura**, Bologna Bo (IT); **Maria Angela Mugnai**, Prato Po (IT)

Correspondence Address:
John Dodds
1707 N St N W
Washington, DC 20036 (US)

(21) Appl. No.: **10/557,426**

(22) PCT Filed: **May 21, 2004**

(86) PCT No.: **PCT/FI04/00310**

§ 371(c)(1),
(2), (4) Date: **Nov. 6, 2006**

(30) **Foreign Application Priority Data**

May 21, 2003 (FI)..... 20030771

Publication Classification

(51) **Int. Cl.**

C12Q 1/68	(2006.01)
C12P 21/06	(2006.01)
C07H 21/04	(2006.01)
C12N 1/21	(2006.01)
C07K 14/195	(2006.01)

(52) **U.S. Cl.** **435/6**; 435/69.1; 435/320.1; 435/252.3; 530/350; 536/23.7

(57) **ABSTRACT**

This invention is related to a method for detecting toxic and non-toxic cyanobacteria. The method comprises that nucleic acid from a biological sample is brought into contact with an oligonucleotide designed to be specific for the mcy gene, in particular mcyE and/or mcyD, and with an oligonucleotide designed to be specific for 16SrDNA, and the presence or absence of toxic cyanobacteria is detected by a suitable molecular biology method. The invention is related also to oligonucleotides used in the method.

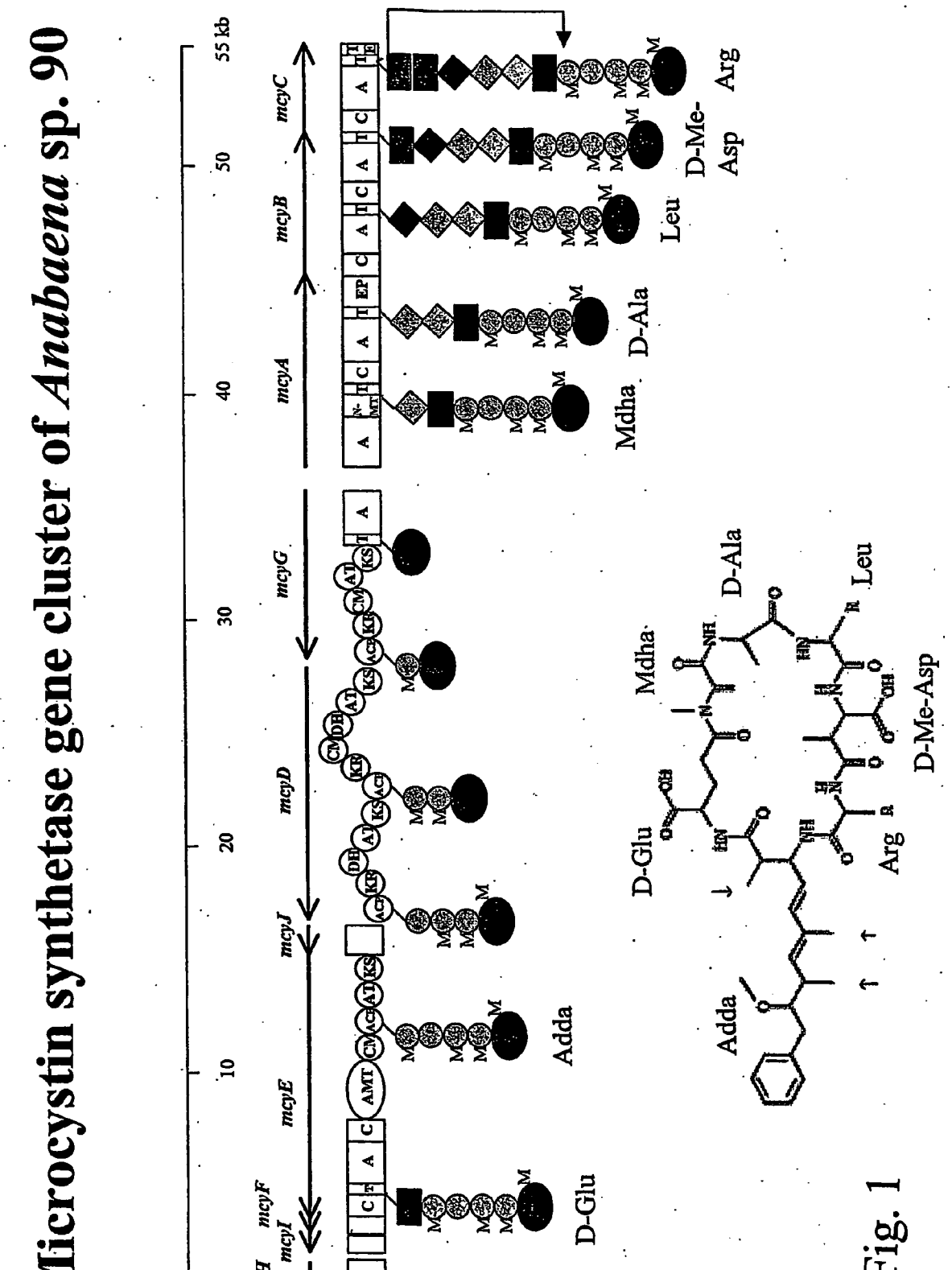
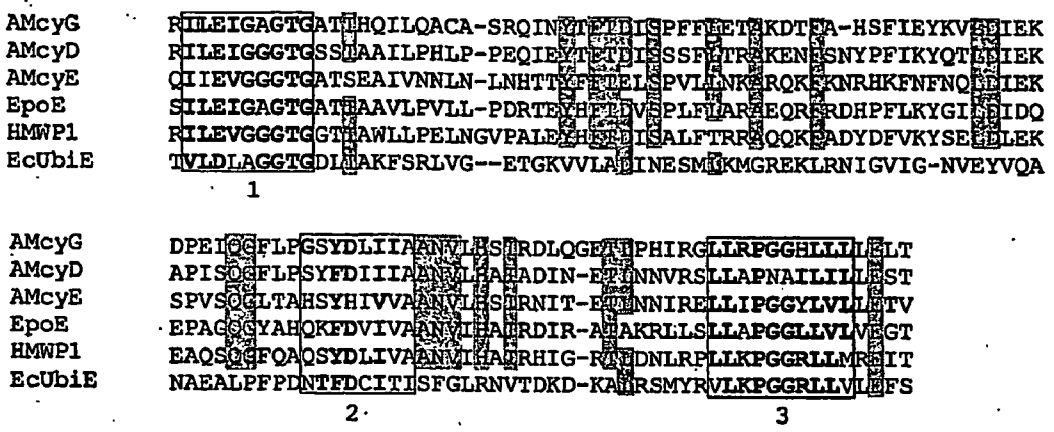

Microcystin synthetase gene cluster of Anabaena sp. 90
0
10
20
30
40
50
55 kb
mcyH
mcyI
mcyF
mcyE
mcyJ
mcyD
mcyG
mcyA
mcyB
mcyC
D-Glu
Adda
Mdha
D-Ala
Leu
D-Me-Asp
Arg

Fig. 1

A


AMcyG
AMcyD
AMcyE
EpoE
HMWP1
EcUbiE
1
AMcyG
AMcyD
AMcyE
EpoE
HMWP1
EcUbiE
2
3

B

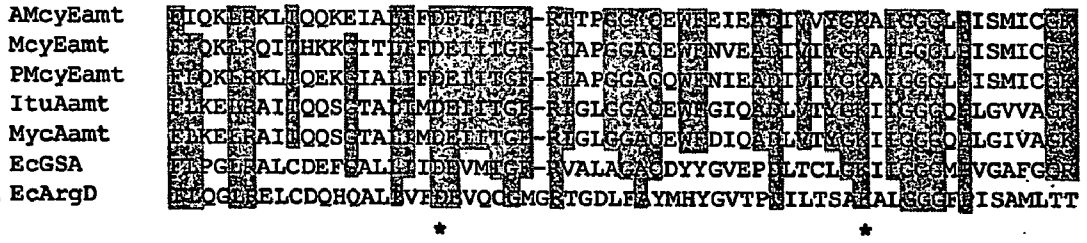

AMcyEamt
McyEamt
PMcyEamt
ItuAamt
MycAamt
EcGSA
EcArgD

Fig. 2

A

```
AMCD-DH2     ASISQNNPEFLTEHQVFDKPIFPGAAFIEMAL
AMCD-DH3     GEISSEYPDYLEGHKVFGKILFPATGFIETIL
RifE-DH10    SRLSLRSHPWLADHAVRDVVIVPGTGLVELAV
RapA-DH4     GRVSLATHAWLADHAVWGRVLLPGTAFVELVV
RapB-DH10    GRVSLATHAWLADHAVRGSVLLPGTGFVELVV
                           *    *     *
```

B

```
AMCG-KR1     QAQATYLITGGIGHLGLQLARHLV-DLGAKHLILTTR
AMCD-KR2     RQDGFYLISGGTGGLGLATARWMI-EHGACHLVLCSR
AMCD-KR3     SKEGAYLITGGLGKLGLLMAQWL-SQMGSSHLVLCSR
RifE-KR10    KTRGPVLVTGGTGSLGGLVARHLVERHGVRQLVLASR
RapA-KR4     DPDGTVLITGGSGVLAGIAARHLVAERGVRHLLLLSR
RapB-KR10    DPDGTVLITGGSGVLAGIAARHLVAERGVRHLLLLSR
                          * *  *   *
```

Fig. 3

AT1	AMcyG	GQGSQY	TSYTQPALFVVEVALAQLW	GHSL	SHAFH	WIQHL
	MMcyG	GQGSQY	TAYTQPALFLIEVALAQLW	GHSL	AHAFH	WLQHL
	PMcyG	GQGSQY	TSYTQPALFIIEVALAQLW	GHSL	SHAFH	WLQHL
AT2	AMcyD	GEVRVN	PLYVHPTLFALQYALCELW	GNGL	WQEFH	WVDAI
	MMcyD	GEAQSH	SLSVQPPLFAYQYALCELW	GSGL	WEAFH	WINSG
	PMcyD	GEAQSN	PLSVQPILFAFHYALCELW	GSGL	WKGFH	WITSV
AT3	AMcyD	GQGSQY	TQITQPALFSLEYALAKLW	GHSI	SHAFH	WLRHL
	MMcyD	GQGSQY	TQITQPVIFSLEYALAKLW	GHSI	SHAFH	WVAHL
	PMcyD	GQGSQY	TQITQPVLFSFEYALAKLW	GHSI	SHAFH	WVNHL
AT4	AMcyE	GQGACY	TAYAQPAIFALEYSLAMLW	GHSV	NGAFH	WRQQS
	MMcyE	GQGACY	TAYAQPAIFALEYSLTMLW	GHSV	TQAFH	WSKQC
	PMcyE	GQGACY	TAYAQPAIFALEYAVAMLW	GHSV	TQAFH	WRQQC
		H	FT E	V L		
Malonyl		GQGXQR	TXYAQXXXXXXQXALXXXX	GHSI	XXAFH	WXXXX
Methymalonyl		GQGXQW	VDVVQXXXXXXMXSLAXXW A	GHSQ *	DYASH	WXXNL

Fig. 4

A

```
AMCG-KS1   GPSVNVQTACSTGLVVVHLA - 122aa - VEAHGTGTKLGDPIE - 18 aa - GSVKTNIGHMQIASGIVGLIK
AMCD-KS2   GPSLAIDTACSSSLVAVHLA - 121aa - VEAHGTGTSLGDPIE - 20 aa - GSVKTNLGHLAAAAGISGLIK
AMCD-KS3   GPSMTVDTACSSSLVAVHLA - 121aa - IEAHGTGTALGDPIE - 19 aa - GSVKTNLGHLEGAAGIAGLIK
AMCE-KS4   GPCMSIDAACASSLVALHQA - 121aa - IEAHGTGTSLGDPIE - 17 aa - GSVKTNIGHLEAAAGIAGTIK
RapA-KS1   GPAITVDTACSSSLVALHQA - 121aa - VEAHGTGTTLGDPIE - 18 aa - GSLKSNIGHAQAAAGVSGVIK
RifA-KS1   GPAVTVDTACSSSLVAMHLA - 121aa - VEAHGTGTTLGDPIE - 18 aa - GSLKSNIGHAQAAAGVAGVIK
                      *
```

B

```
AMCG-ACP1  EQGFLEMGIDSLLSIELKNRLEKGLEVALPASLIF
AMCD-ACP2  QQGFFDMGMDSLTSTELRNLLQTDFNCSLPTTIAF
AMCD-ACP3  HTSFLELGLNSLMVLEFKNRLQSNLACTLPTSIIF
AMCE-ACP4  DETLLNLGADSIILTDFVRKIEEKFGVKVKIDQLF
RapA-ACP1  TGAFRDLGVDSLTAVELRNGLAKATGLRLPATLVF
RapC-ACP11 TTAFKDLGINSLTAVELRNSLAKATELRLPATLVF
RifA-ACP1  GRTFKDAGFDSLTAVELRNRLAAATGLTLSPAMIF
                     *
```

Fig. 5

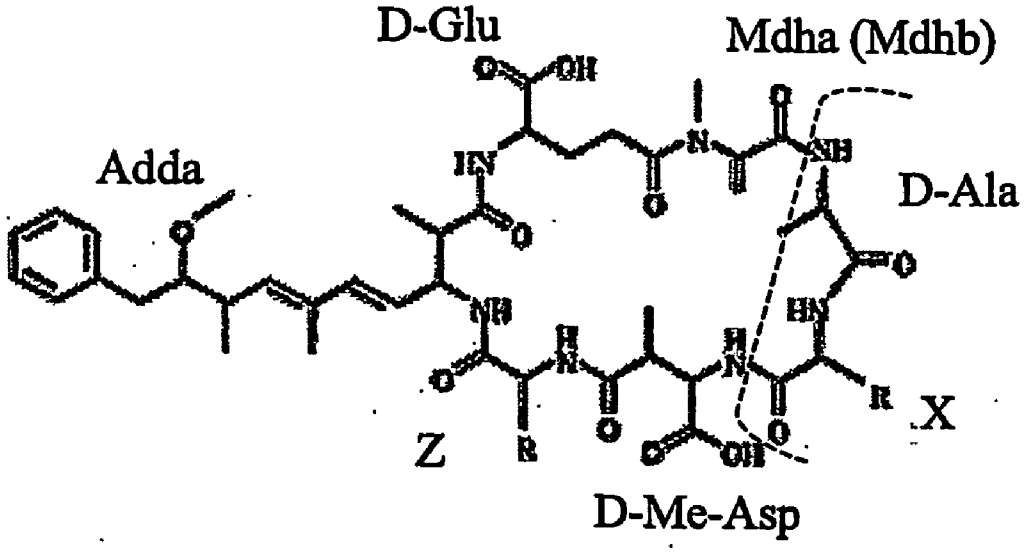


Fig. 6

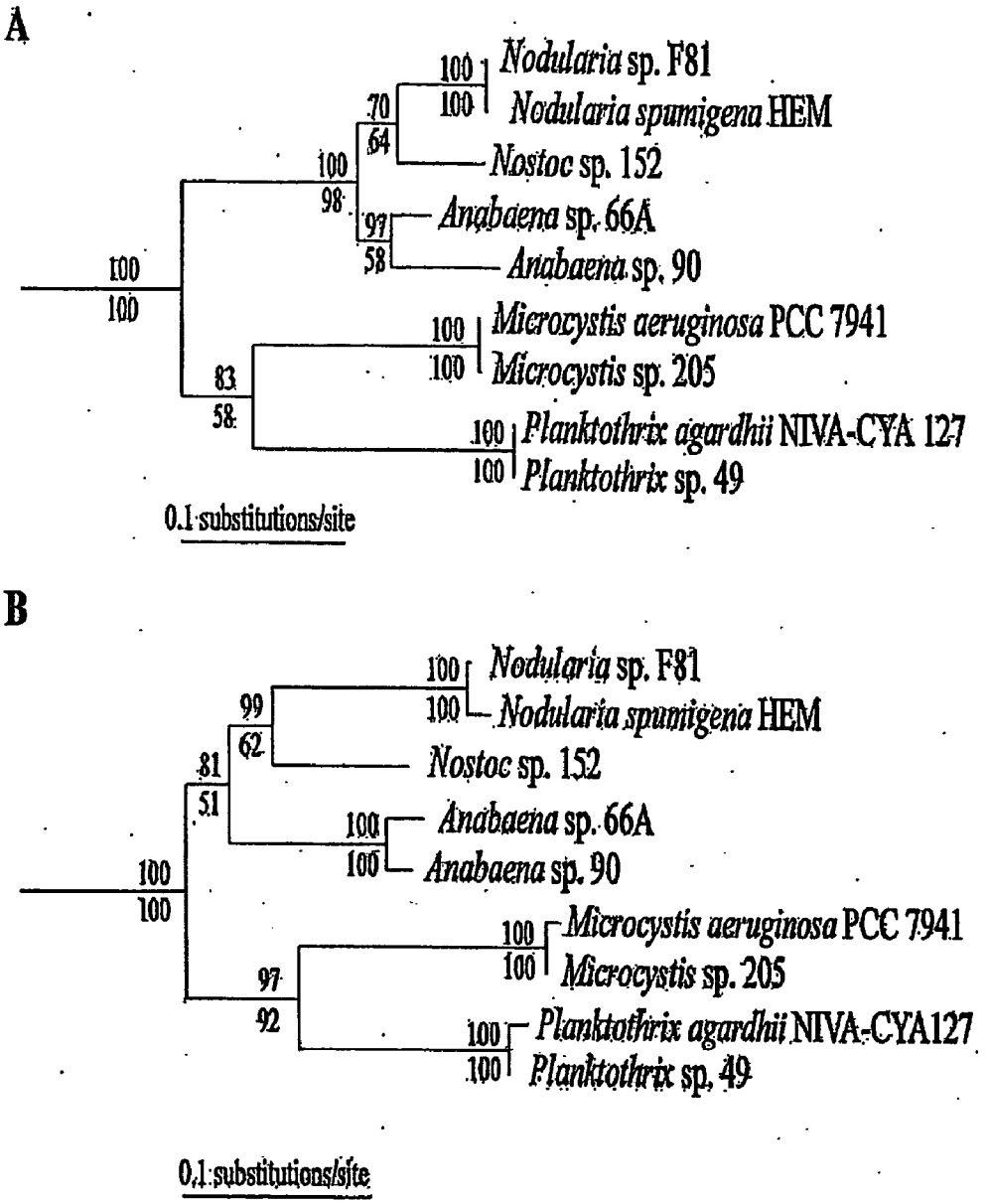

A
100
100
Nodularia sp. F81
70
64
Nodularia spumigena HEM
100
98
Nostoc sp. 152
97
58
Anabaena sp. 66A
100
100
Anabaena sp. 90
100
100
Microcystis aeruginosa PCC 7941
83
58
Microcystis sp. 205
100
100
Planktothrix agardhii NIVA-CYA 127
Planktothrix sp. 49
0.1 substitutions/site
B
100
100
Nodularia sp. F81
99
62
Nodularia spumigena HEM
81
51
Nostoc sp. 152
100
100
Anabaena sp. 66A
100
100
Anabaena sp. 90
100
100
Microcystis aeruginosa PCC 7941
97
92
Microcystis sp. 205
100
100
Planktothrix agardhii NIVA-CYA127
Planktothrix sp. 49
0.1 substitutions/site

Fig. 7

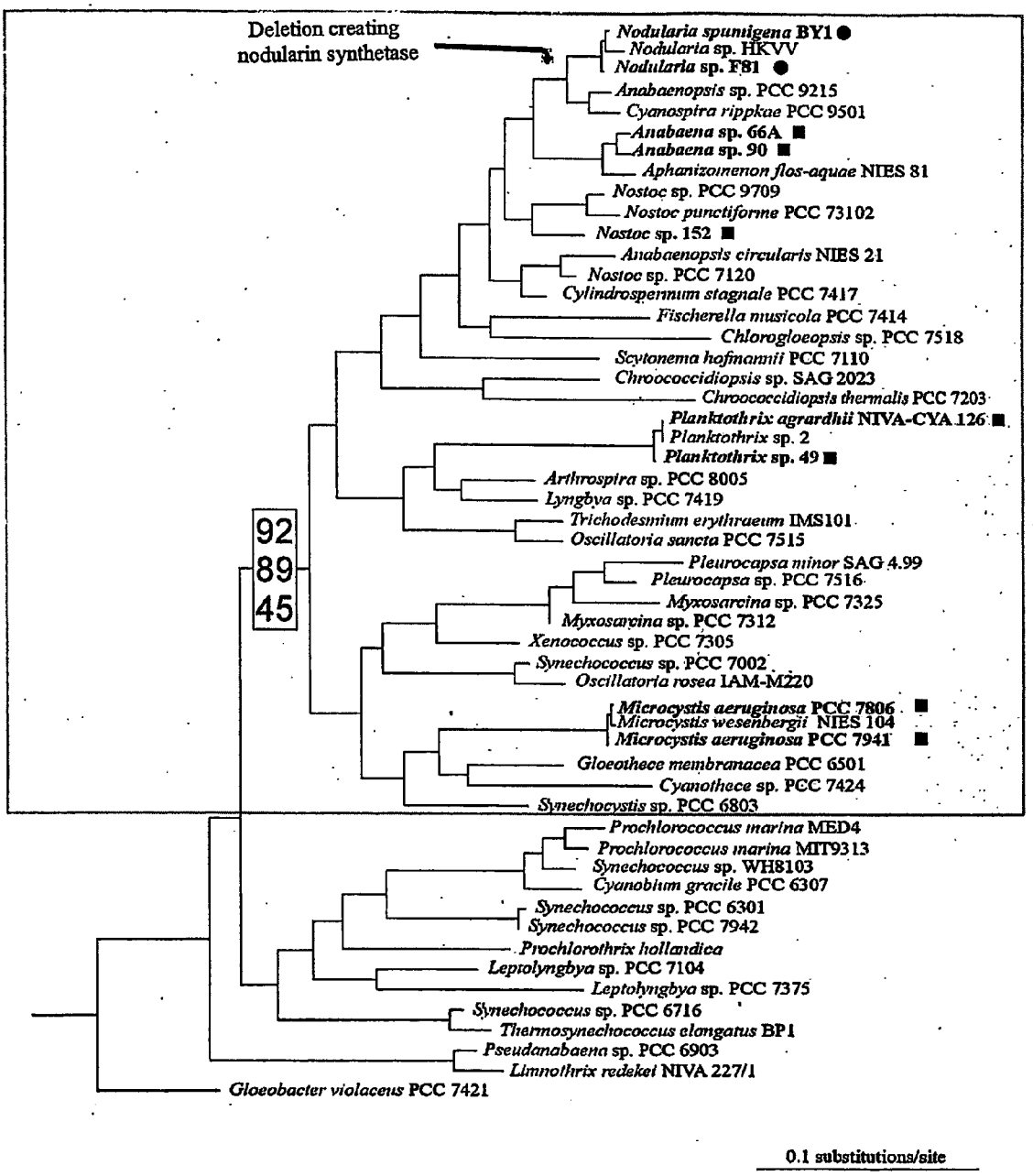
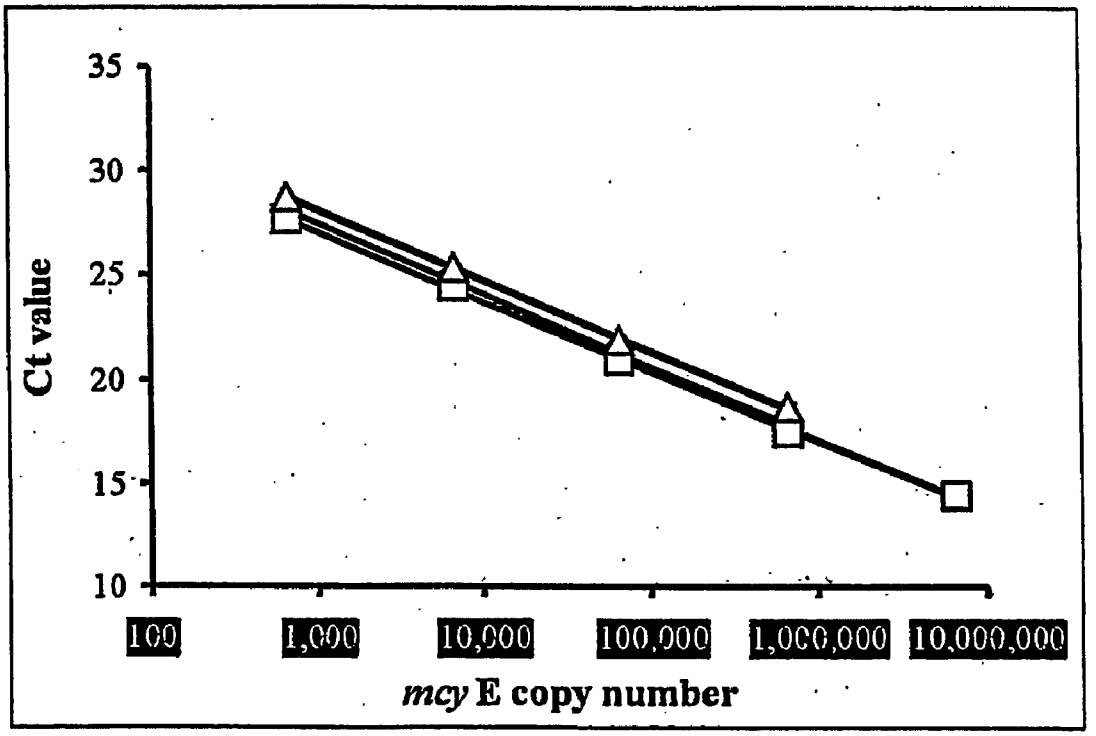

Deletion creating nodularin synthetase
Nodularia spumigena BY1
Nodularia sp. HKVV
Nodularia sp. F81
Anabaenopsis sp. PCC 9215
Cyanospira rippkae PCC 9501
Anabaena sp. 66A
Anabaena sp. 90
Aphanizomenon flos-aquae NIES 81
Nostoc sp. PCC 9709
Nostoc punctiforme PCC 73102
Nostoc sp. 152
Anabaenopsis circularis NIES 21
Nostoc sp. PCC 7120
Cylindrospermum stagnale PCC 7417
Fischerella musicola PCC 7414
Chlorogloeopsis sp. PCC 7518
Scytonema hofmannii PCC 7110
Chroococcidiopsis sp. SAG 2023
Chroococcidiopsis thermalis PCC 7203
Planktothrix agrardhii NIVA-CYA 126
Planktothrix sp. 2
Planktothrix sp. 49
Arthrospira sp. PCC 8005
Lyngbya sp. PCC 7419
Trichodesmium erythraeum IMS101
Oscillatoria sancta PCC 7515
92
89
45
Pleurocapsa minor SAG 4.99
Pleurocapsa sp. PCC 7516
Myxosarcina sp. PCC 7325
Myxosarcina sp. PCC 7312
Xenococcus sp. PCC 7305
Synechococcus sp. PCC 7002
Oscillatoria rosea IAM-M220
Microcystis aeruginosa PCC 7806
Microcystis wesenbergii NIES 104
Microcystis aeruginosa PCC 7941
Gloeothece membranacea PCC 6501
Cyanothece sp. PCC 7424
Synechocystis sp. PCC 6803
Prochlorococcus marina MED4
Prochlorococcus marina MIT9313
Synechococcus sp. WH8103
Cyanobium gracile PCC 6307
Synechococcus sp. PCC 6301
Synechococcus sp. PCC 7942
Prochlorothrix hollandica
Leptolyngbya sp. PCC 7104
Leptolyngbya sp. PCC 7375
Synechococcus sp. PCC 6716
Thermosynechococcus elongatus BP1
Pseudanabaena sp. PCC 6903
Limnothrix redekei NIVA 227/1
Gloeobacter violaceus PCC 7421
0.1 substitutions/site

Fig. 8


35
30
25
20
15
10
Ct value
100
1,000
10,000
100,000
1,000,000
10,000,000
mcy E copy number

A

Fig. 9A

B

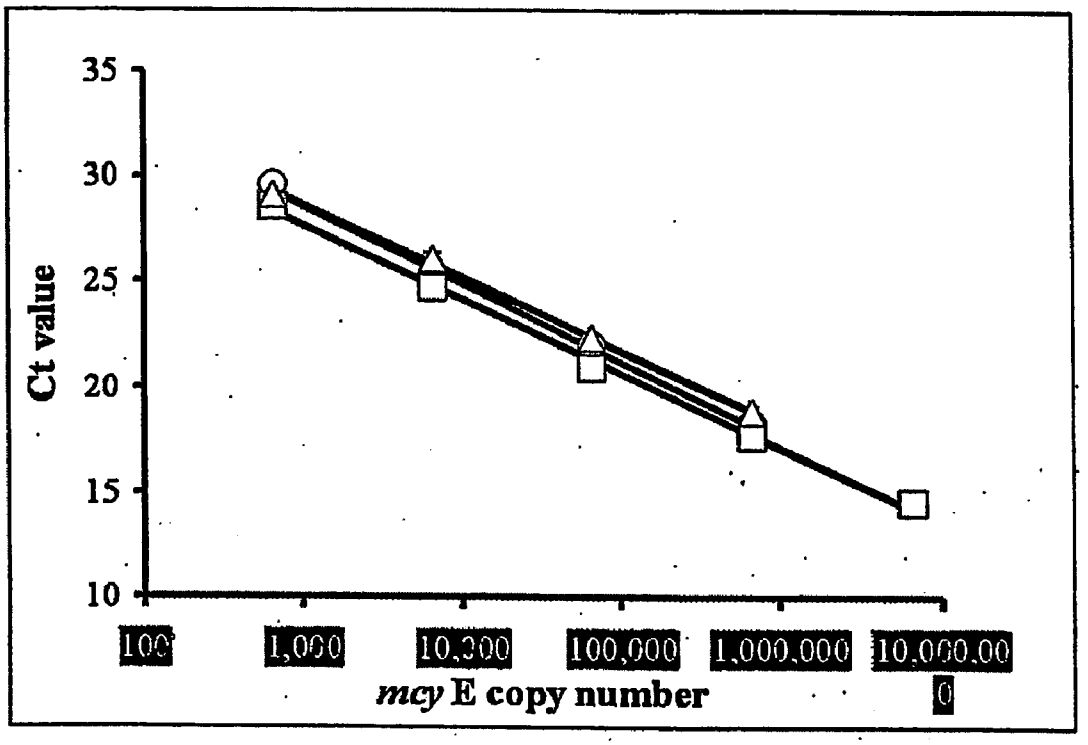

35
30
25
20
15
10
Ct value
100
1,000
10,000
100,000
1,000,000
10,000,00
0
mcy E copy number

Fig. 9B

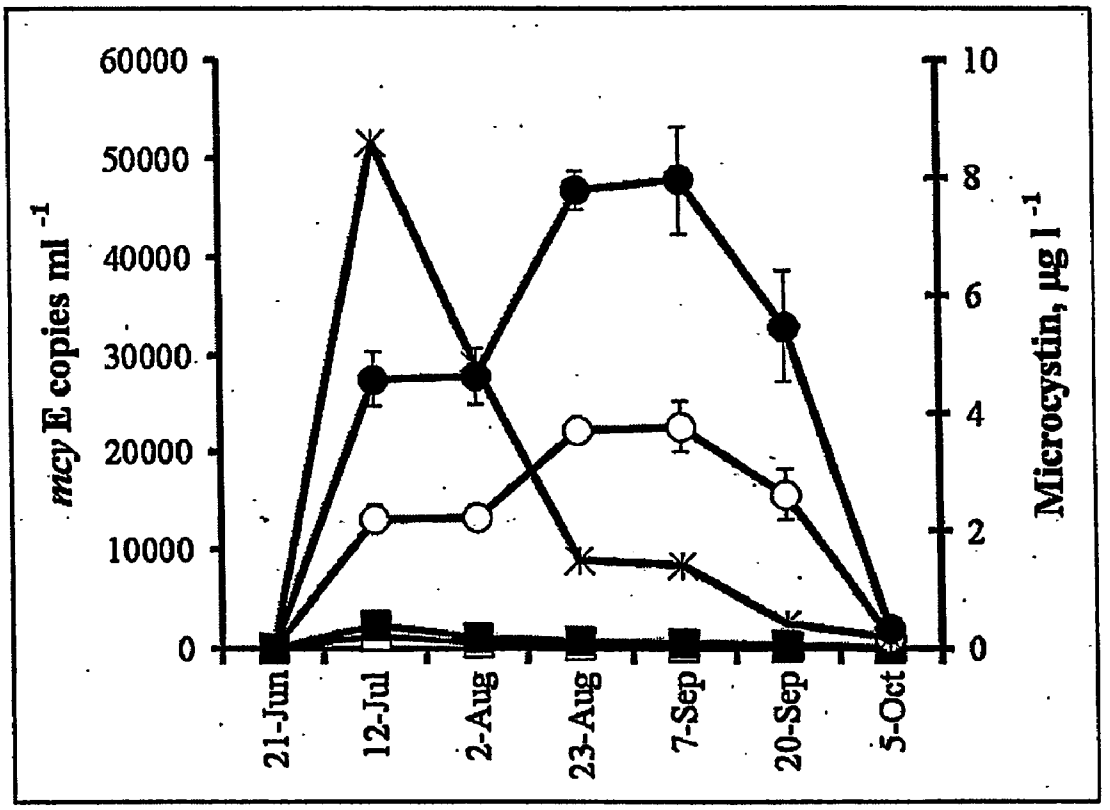

60000
50000
40000
30000
20000
10000
0
mcy E copies ml -1
10
8
6
4
2
0
Microcystin, µg l -1
21-Jun
12-Jul
2-Aug
23-Aug
7-Sep
20-Sep
5-Oct

Fig. 10

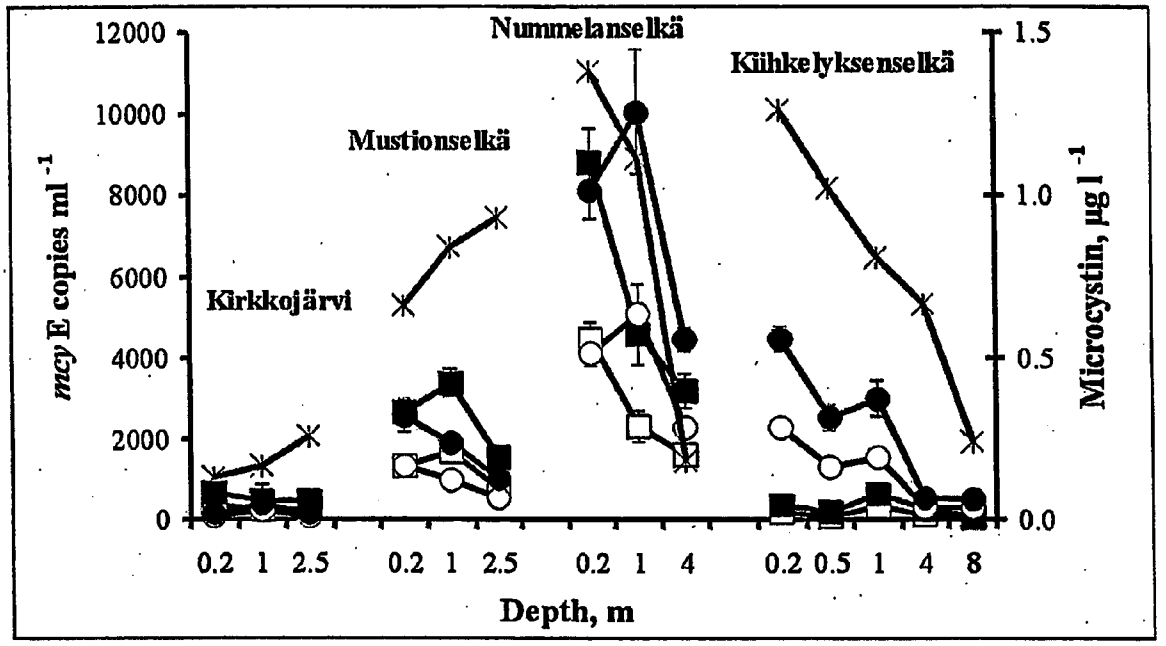

Nummelanselkä
Kiihkelyksenselkä
Mustionselkä
Kirkkojärvi
12000
10000
8000
6000
4000
2000
0
1.5
1.0
0.5
0.0
mcyE copies ml -1
Microcystin, µg l -1
0.2 1 2.5
0.2 1 2.5
0.2 1 4
0.2 0.5 1 4 8
Depth, m

Fig. 11

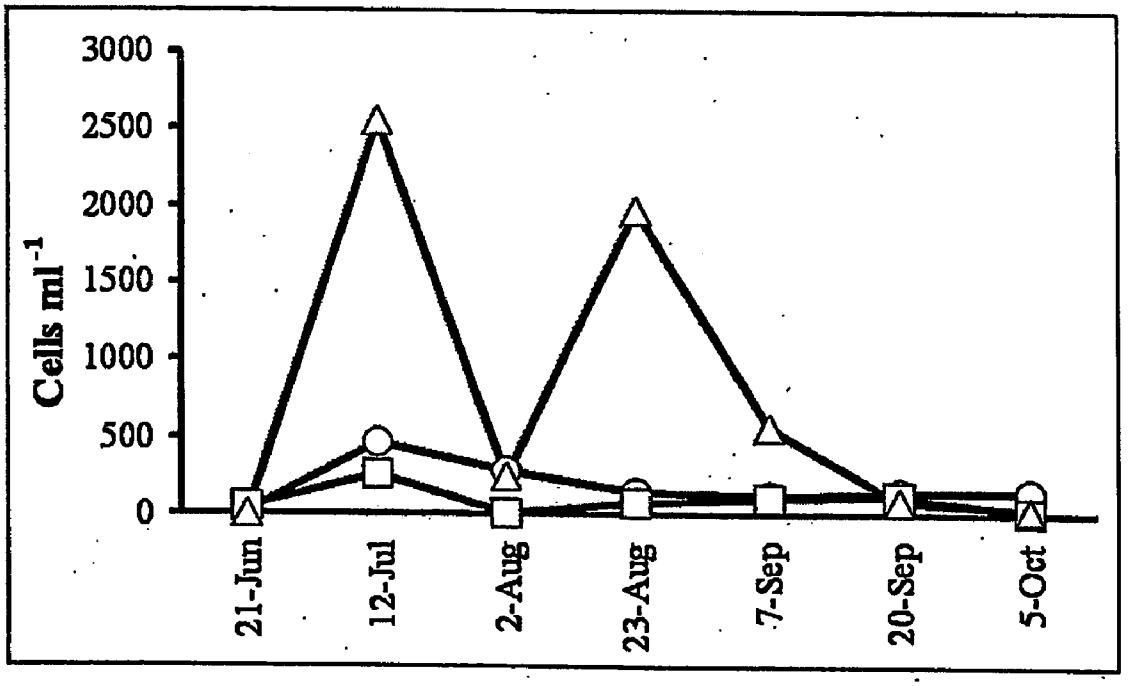

3000
2500
2000
1500
1000
500
0
Cells ml-1
21-Jun
12-Jul
2-Aug
23-Aug
7-Sep
20-Sep
5-Oct

Fig. 12

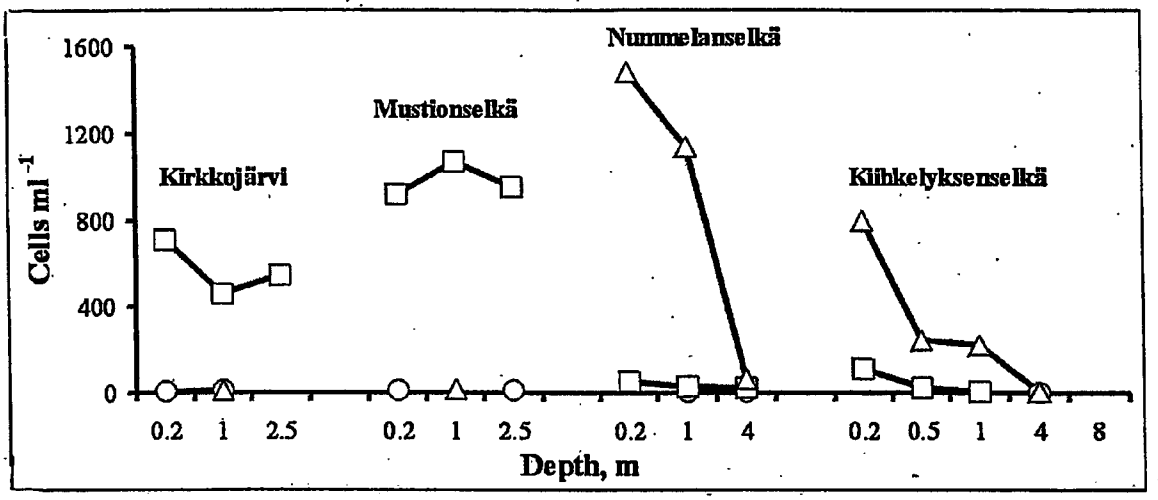

Nummelanselkä
Mustionselkä
Kirkkojärvi
Kiihkelyksenselkä
Cells ml -1
1600
1200
800
400
0
0.2 1 2.5
0.2 1 2.5
0.2 1 4
0.2 0.5 1 4 8
Depth, m

Fig. 13

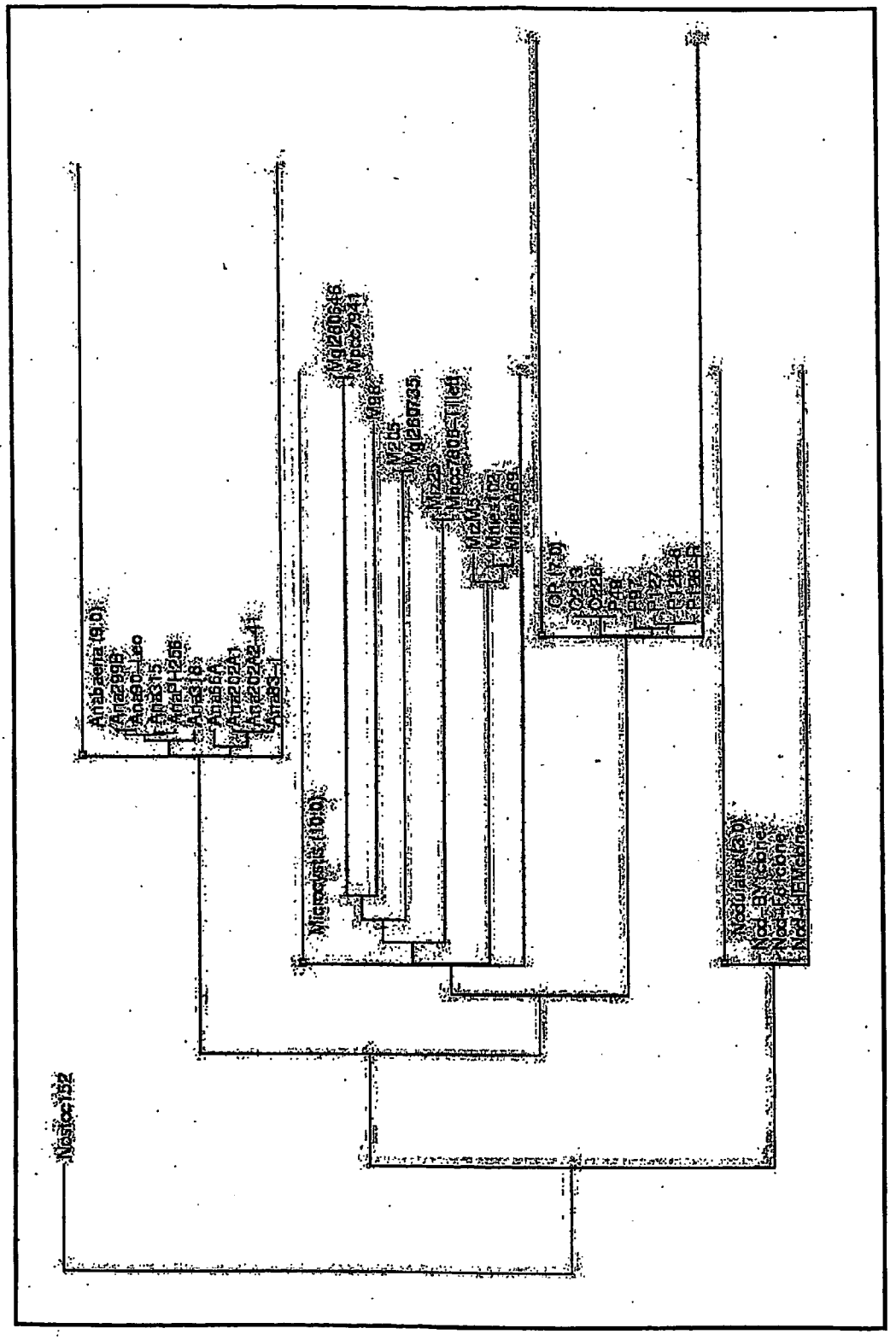

Anabaena (9,0)
Microcystis (10,0)
Nostoc/152
Nostoc/Peltigera
Nostoc/Blasia
Nostoc/Anthoceros
Nostoc/Gunnera
Nostoc/Cycas
Nostoc/Blasia
Nostoc/Peltigera
Nostoc/Anthoceros
Nostoc/Gunnera
Nostoc/Cycas

Fig. 14

```
                                    1                                                  50
Consensus_Anabaena_McyE             AGCATTAGCA GTTGGTTATC GTAATCTTCC TGAAATTACC AGAGAAAAAT
Consensus_Microcystis_McyE          CGCCCTAGCA TCGGGTTATC ATAACCAACC CGAAATGACT CAAGAAAAAT
Consensus_Nodularia_McyE            AGCATTAGCA KCAGGTTATC ACAACCTCCC CCAAATCACC GCAGAAAAAT
Consensus_Nostoc_McyE               AGCATTAGCC GCAGGTTATC ATAATCTTCC CGACATCACT ACAGAAAAAT
Consensus_Oscillatoria/Planktothrix_McyE AGCCTTAGCA TCGGGTTATC ACAATCTTCC TCAAATCACA AAAGAAAAAT

                                    51                                                 100
Consensus_Anabaena_McyE             TTCAACCCAA CTTTTTAAAT TCAGAAAAAA TTCTCTTTAG AACGGGAGAT
Consensus_Microcystis_McyE          TTAAACCTAG CTTTCTTGAT GAGACAAAAA CTCTCTTTAG AACCGGCGAT
Consensus_Nodularia_McyE            TTCAACCTAG CTYTATGACT GAGGGAAAAA CTATCTTTAG AACCGGAGAT
Consensus_Nostoc_McyE               TTCAACCCAG CTTGATAAGT GAGGGAAAAA CTCTCTTTAG AACGGGAGAT
Consensus_Oscillatoria/Planktothrix_McyE TTAAACCTGG CTTTTTTRAT CAGAAAACAA CGATGTTTAG AACCGGGGAT

                                    101                                                150
Consensus_Anabaena_McyE             TTAGGTAAAC AAATTGCTCC AGGTGTGATT GAATTTATAG GGCGAAAAGA
Consensus_Microcystis_McyE          TTAGGCAAGC AAACTGCTCC SGGTATCATT GAGTTTATGG GACGAAAAGA
Consensus_Nodularia_McyE            TTAGGTAAAC AAATTGCCCC AGGCGTGATT GAATTTCTTG GTCGTAAAGA
Consensus_Nostoc_McyE               TTAGGTAAAC AAACTGCTCC AGGTGTCATT GAATTTATTG GGCGTAAAGA
Consensus_Oscillatoria/Planktothrix_McyE TTAGGCAAAC AAACTGCTCC CGGTGTGATT GAATTTATGG GCAGAAAAGA

                                    151                                                200
Consensus_Anabaena_McyE             TAATCAAGTT AAGGTCAATG GCTATCGTGT AGATCCAGGA GAAMTTGAAT
Consensus_Microcystis_McyE          TAATCAAGTT AAGGTCAATG GTTATCGAAT TGACCCCGGA GAAATTGAAT
Consensus_Nodularia_McyE            TAATCAAGTT AAGGTGAATG GCTATCGYAT AGATCCAGGA GAAATTGAAT
Consensus_Nostoc_McyE               TAATCAAGTA AAGGTGAATG GTTATCGTAT AGATCCAGGA GAAATTGAAT
Consensus_Oscillatoria/Planktothrix_McyE CAATCAAGTT AAGGTAAATG GCTATCGTAT CGACCCCGAA GAAATTGAAT

                                    201                                                250
Consensus_Anabaena_McyE             ACCAAATTAG CCGCTATGCC TGAGATTGAGA AAGCAATTGT CTTACCTATA
Consensus_Microcystis_McyE          ATCAATTGAC TCGTTATGCT CCCATTGAAA GAGCGATTGT TTTACCCGTT
Consensus_Nodularia_McyE            ACCAACTCAG CCGCCATTCT CAAATTGAGA GAGCAATCGT ATTGCCTACT
Consensus_Nostoc_McyE               ATCAACTCAG CCGTCATGCT GAGATTGAAA GAGCGATTAT ATTGCCTATC
Consensus_Oscillatoria/Planktothrix_McyE ATCAACTTAA TCGTTATCCT GAGATTGAGA GAGCAATTAT TCTACCGATA

                                    251                                                300
Consensus_Anabaena_McyE             GAGGTAAATA ACCAAATTCA ATTATCTGCT TATTGTCAAA CTGATAAAGA
Consensus_Microcystis_McyE          CAAGTKAATA ATCAAACTCA ATTATCTGCT TACTGTCAAA CAGACAAAAC
Consensus_Nodularia_McyE            AATGTAGATA ATCAAACCCA GTTATCAGCC TATTGTAAAA CTGAGTCAGA
Consensus_Nostoc_McyE               AATGTAGATA ATCAAACTCA ATTATCTGCT TATTGTCAAA CTGATAAAGA
Consensus_Oscillatoria/Planktothrix_McyE TCAGTCAATA ATCAAACTCA ATTATCAGCC TATTGTCAAA CMGATAAACA
```

Fig. 15A

```
                                          301                                                 350
Consensus_Anabaena_McyE                   TATAAAAWTT TCTGAAATCA GAGAATTTTT AGCTAAATAT TYGCCAGTTT
Consensus_Microcystis_McyE                TCTAGAAATT GCTGAGATTC GAGAATTWCT TGCCAARTTT TTACCAGTTT
Consensus_Nodularia_McyE                  CATAGAAATT TCCGAAATTC GAGAATTTCT ATCGAACTTT TTGCCCGTTT
Consensus_Nostoc_McyE                     CATCGAAATT GCTGAAATTA GAGAATTTCT CTCTAAATTT TTGCCAGTTT
Consensus_Oscillatoria/Planktothrix_McyE  GATAGAAATT TCTGAAATYA GAGAATTTCT AGCTAATTTT CTGCCAGTTT

                                          351                                                 400
Consensus_Anabaena_McyE                   ACATGATTCC TAGTTCCTTT ATCTTCTTAA AGCAATTTCC CTTAACTAAA
Consensus_Microcystis_McyE                ATATGATTCC SAGTTACTTT ATTTTTTTAA AGCAATTCCC CTTAACTCGA
Consensus_Nodularia_McyE                  AYATGATTCC TACTTTCTTT ATCTTCTTAA AGCAATTTCC CTTAACCAGA
Consensus_Nostoc_McyE                     ATATGATTCC TACTTCCTTT ATCTTCCTCA AGCAATTTCC TCTAACCAGA
Consensus_Oscillatoria/Planktothrix_McyE  ACATGATTCC TAGTTACTTT ATTTTCTTAA AGCAATTYCC CCTAACTAAA

                                          401                                                 450
Consensus_Anabaena_McyE                   CATGGCAAAY TTGACTTGCG ATCGCTCRTC GCTGTCAAGC CRACAGATCA
Consensus_Microcystis_McyE                CATGGRAAAC TTGACCTGCA CTCCCTGAGA GAACTCAGAG AAACYGGTAA
Consensus_Nodularia_McyE                  CATGGGAAAA TTGATTTGCG ATCCCTGGCT GAAATCAAGG GAATAGGTAA
Consensus_Nostoc_McyE                     CATGGCAAAA TTGACTTGCG ATCGCTTGCT GAAGTCCAAG GAATCGGTAA
Consensus_Oscillatoria/Planktothrix_McyE  CACGGCAAAC TTGACTTAAA CTCAMTGATT GCAGTCAATG AAACCGGGAA

                                          451                                                 500
Consensus_Anabaena_McyE                   AY...TAACA CAAGTCTCTT ATACTGCACC GCGTAATACT TTAGAATCAA
Consensus_Microcystis_McyE                ATCTCTGGTK AATTCTAATT AYGTTGCACC CCGKAATYAT TTAGAATCCA
Consensus_Nodularia_McyE                  CT...TAACA CAGTTAGCGT ATACTGCACC GCGCAATAAT TTAGAGTCCA
Consensus_Nostoc_McyE                     GT...TAACA CAGGCAGATT ATACTGCACC GCGCAATGAT TTAGAGTCCA
Consensus_Oscillatoria/Planktothrix_McyE  AT...CTACC CARGTAAATT ATGTTGCACC GCGTAATAAT TTAGAGTCAA

                                          501                                                 550
Consensus_Anabaena_McyE                   AGCTAGTCCA TATTTGGGAA AAAATTCTCA CTAAACATCC CATTGGAATT
Consensus_Microcystis_McyE                ATCTCGTTAG TATCTGGGAA AAAATTCTCT CTAAACATCC TATCGGTATT
Consensus_Nodularia_McyE                  AGCTCGTACA TATTTGGGAA AAAATTCTCA CCAAACAACC CATTGGCATT
Consensus_Nostoc_McyE                     AGCTAGTAAA GATTTGGGAA AAAATTCTCA CCACACATCC CATCGGCATT
Consensus_Oscillatoria/Planktothrix_McyE  ACCTAGTTAG AATCTGGGAA AAGATTCTRA CCAAACATCC CATCGGTATT

                                          551                                                 600
Consensus_Anabaena_McyE                   TTTGATAACT TTTTTGAAAT CGGCGGACAC TCTCTGCTCC TTTCTAGAGT
Consensus_Microcystis_McyE                TTTGATAACT TCTTTGAAAT TGGCGGTCAT TCTCTACTCT TATCAAGGGT
Consensus_Nodularia_McyE                  TTTGATAACT TCTTTGAAAT TGGTGGACAC TCACTGCTGC TTTCCAGAGT
Consensus_Nostoc_McyE                     TTTGATAACT TCTTTGAAAT TGGTGGACAC TCSCTGCTGC TTTCGAGAGT
Consensus_Oscillatoria/Planktothrix_McyE  TTTGATAACT TCTTTGAAAT TGGCGGACAT TCTCTGATGC TTTCGAGAAT
```

Fig. 15B

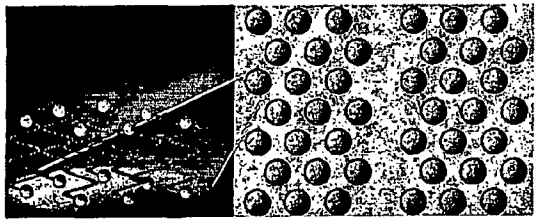
```
                                          601                                                650
Consensus_Anabaena_McyE                   AGTCACTCAG GTCCATAAAG AATTAAATGT ATTAGTTAAA TTAGCTGATT
Consensus_Microcystis_McyE                TGTAACCCGG GTTCATAAAG AACTAAATGT ATCCGTAAAA TTRGCTGACT
Consensus_Nodularia_McyE                  GGTAACTCAC GTTCATAAAG AATTAAATGT GTTGGTAAAA TTGGCTGAAT
Consensus_Nostoc_McyE                     GGTAACTTAG GTTCATAAAG AATTAAATGT GTTAGTCAAA TTGGCTGACT
Consensus_Oscillatoria/Planktothrix_McyE  CGTAACCCAG GTTCATAAAG AATTAAATGT ATCGGTAAAA TTGGCTGACT

                                          651                                                700
Consensus_Anabaena_McyE                   TCTTCAAAGT TCCCACAATT CTTGGATTAG CAGCTTTAAT ATCTAAAGCT
Consensus_Microcystis_McyE                TCTTTAAAGT TCCAACCRTT GCTGGATTGG CGACTTTAAT CTCCCAGACT
Consensus_Nodularia_McyE                  TCTTTAAAGT TCCCACAATC GCCGGATTAG CAGCTTTAGT ATCTAAAACC
Consensus_Nostoc_McyE                     TTTTTAAAGT TCCCACCATA GCCGGATTAG CAGCTTTAGT AGCTAAAACC
Consensus_Oscillatoria/Planktothrix_McyE  TCTTTAAAGT TCCTACCATT GCCGGATTAG CCGTTTTAGT CTCTAAAACT

                                          701                                                750
Consensus_Anabaena_McyE                   CAATCTAACT ATCAAGAACC CATACCAGCA ATAACTCAAC AAGAATCTTA
Consensus_Microcystis_McyE                CAATACAATT ATCAAGAACC CATTTCGGCA ATTCCCCCCC AAAAATCYTA
Consensus_Nodularia_McyE                  CAATATGACT ATCAAGAACC CATACCAGCA ATAACTCAGC AAACGTCTTA
Consensus_Nostoc_McyE                     CAATACGATT ATCAAGAACC CATACCTGCA ATAATTCAGC AAAAATCTTA
Consensus_Oscillatoria/Planktothrix_McyE  GAATATGATT ATCAAGAACC CATYCCCACA ATTCCTCTGC AAAAATCCTA

                                          751                772
Consensus_Anabaena_McyE                   TCCCATGTCT CATGGACAAC GC  (SEQ ID NO:35)
Consensus_Microcystis_McyE                TCCGATGTCT CATGGTCAGC GT  (SEQ ID NO:36)
Consensus_Nodularia_McyE                  TCCTATGTCT CATGGGCAAC GC  (SEQ ID NO:37)
Consensus_Nostoc_McyE                     TCCCATGTCT CATGGGCAAC GC  (SEQ ID NO:38)
Consensus_Oscillatoria/Planktothrix_McyE  TCCGATGTCC CATGGGCAAC GT  (SEQ ID NO:39)
```

Fig. 15C

Fig. 16

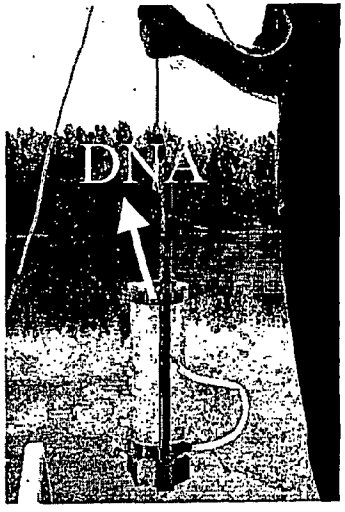
DNA-chip (Microarray)

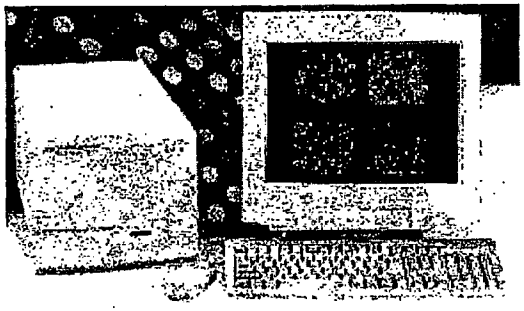
EU project "MIDI-CHIP"

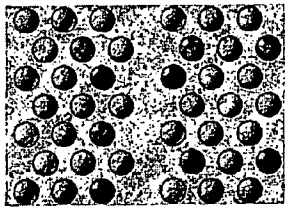


- Probes on a DNA-chip
- large amounts of sequence data is needed for probe design

DNA extracted from environmental samples (PCR product) reacts with the probes on the chips







Chips designed by CNR-ITBA

Chips are scanned with laser and results analyzed with computer

- genetic identification of cyanobacteria present in the sample (16S rRNA)
- presence of cyanobacteria with microcystin synthetase genes (*mcy*E)

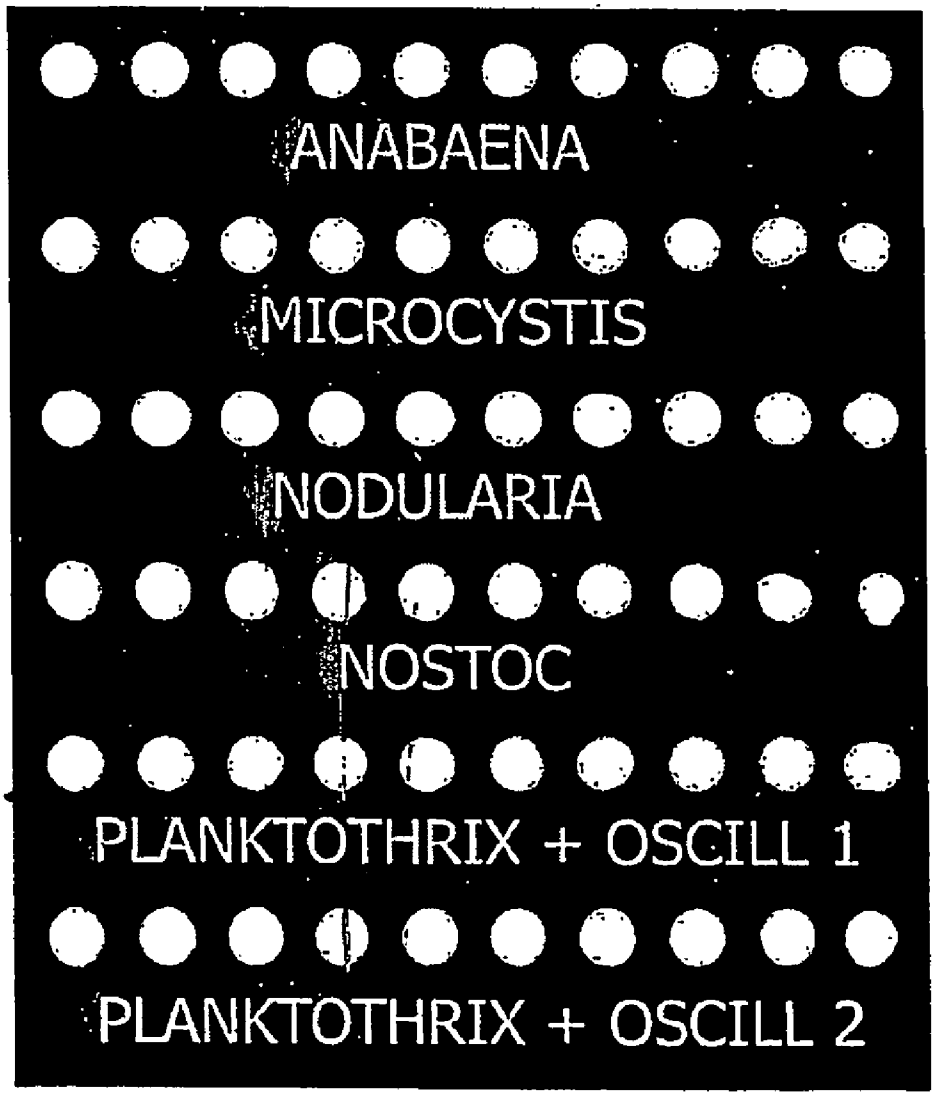
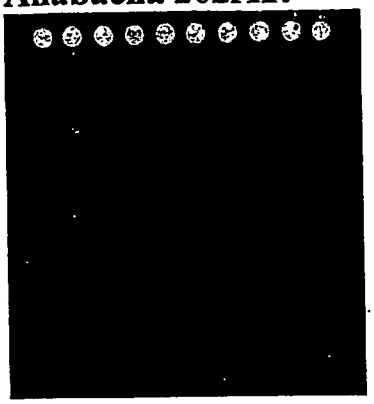
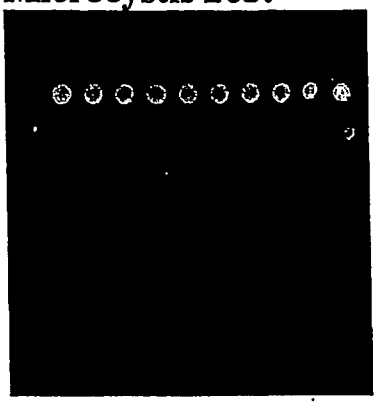

ANABAENA
MICROCYSTIS
NODULARIA
NOSTOC
PLANKTOTHRIX + OSCILL 1
PLANKTOTHRIX + OSCILL 2

Fig. 17

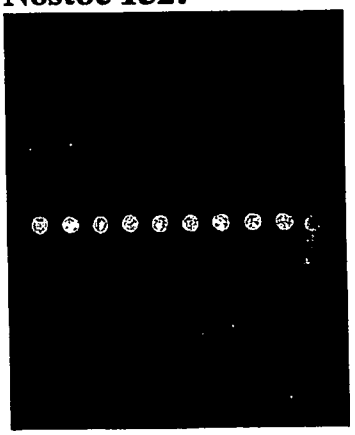
Anabaena 202A1:



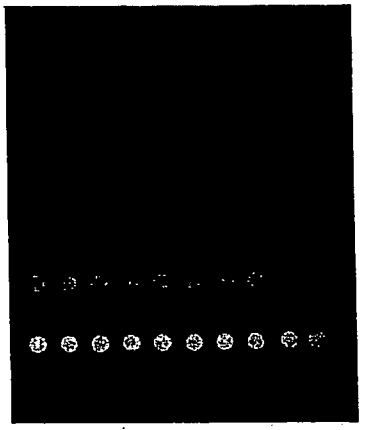
Microcystis 205:



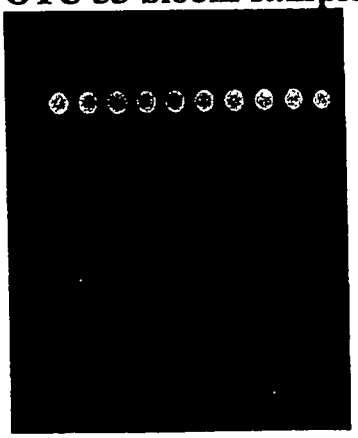
Nostoc 152:



Planktothrix 49:



OTU 33 bloom sample:



OTU 35 >10 um fraction:

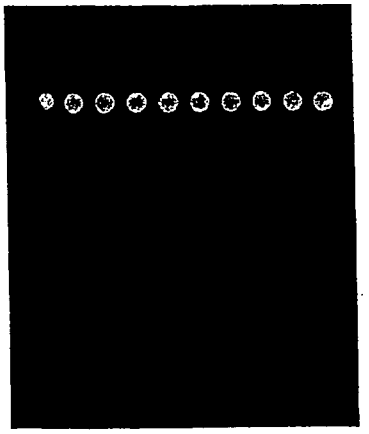


Fig. 18

Fig. 19A

Alignment

```
                    ....|....|  ....|....|  ....|....|  ....|....|  ....|....|  ....|....|
                         5          15          25          35          45          55
Ana90-Leo           AGCATTAGCA GTTGGTTATC GTAATCTTCC TGAAATTACC AGAGAAAAAT TTCAACCCAA
Ana299B             AGCATTAGCA GTTGGTTATC GTAATCTTCC TGAAATTACC AGAGAAAAAT TTCAACCCAA
Ana315              AGCATTAGCA GTTGGTTATC GTAATCTTCC TGAAATTACC AGAGAAAAAT TTCAACCCAA
AnaPH256            AGCATTAGCA GTTGGTTATC GTAATCTTCC TGAAATTACC AGAGAAAAAT TTCAACCCAA
Ana318              AGCATTAGCA GTTGGTTATC GTAATCTTCC TGAAATTACC AGAGAAAAAT TTCAACCCAA
Ana202A1            AGCATTAGCA GTTGGTTATC GTAATCTTCC TGAAATTACC AGAGAAAAAT TTCAACCCAA
Ana202A2-41         AGCATTAGCA GTTGGTTATC GTAATCTTCC TGAAATTACC AGAGAAAAAT TTCAACCCAA
Ana66A              AGcATTAGCA GTTGGTTATC GTAATCTTCC TGAAATTACC AGAGAAAAAT TTCAACCCAA
Ana83-1             AGCATTAGCA GTTGGTTATC GTAATCTTCC TGAAATTACC AGAGAAAAAT TTCAACCCAA
Cons-Ana98%         AGCATTAGCA GTTGGTTATC GTAATCTTCC TGAAATTACC AGAGAAAAAT TTCAACCCAA
Mpcc7806-Tillet     CGCCCTAGCA TCGGGTTATC ATAACCAACC CGAAATGACT CAAGAAAAAT TTAAACCTAG
M205                CGCCCTAGCA TCGGGTTATC ATAACCAACC CGAAATGACT CAAGAAAAAT TTAAACCTAG
M98                 CGCCCTAGCA TCGGGTTATC ATAACCAACC CGAAATGACT CAAGAAAAAT TTAAACCTAG
Mgl260735           CGCCCTAGCA TCGGGTTATC ATAACCAACC CGAAATGACT CAAGAAAAAT TTAAACCTAG
Mgl280646           CGCCCTAGCA TCGGGTTATC ATAACCAACC CGAAATGACT CAAGAAAAAT TTAAACCTAG
Mpcc7941            CGCCCTAGCA TCGGGTTATC ATAACCAACC CGAAATGACT CAAGAAAAAT TTAAACCTAG
Miz25               CGCCCTAGCA TCGGGTTATC ATAACCAACC CGAAATGACT CAAGAAAAAT TTAAACCTAG
MizM5               CGCCCTAGCA TCGGGTTATC ATAACCAACC CGAAATGACT CAAGAAAAAT TTAAACCTAG
Mnies102            CGCCCTAGCA TCGGGTTATC ATAACCAACC CGAAATGACT CAAGAAAAAT TTAAACCTAG
MniesA89            CGCCCTAGCA TCGGGTTATC ATAACCAACC CGAAATGACT CAAGAAAAAT TTAAACCTAG
Cons-Mic97,4%       CGCCCTAGCA TCGGGTTATC ATAACCAACC CGAAATGACT CAAGAAAAAT TTAAACCTAG
P49                 AGCCTTAGCA TCGGGTTATC ACAATCTTCC TCAAATCACA AAAGAAAAAT TTAAACCTGG
P97                 AGCCTTAGCA TCGGGTTATC ACAATCTTCC TCAAATCACA AAAGAAAAAT TTAAACCTGG
P126-8              AGCCTTAGCA TCGGGTTATC ACAATCTTCC TCAAATCACA AAAGAAAAAT TTAAACCTGG
P127                AGCCTTAGCA TCggGTTATC ACAATCTTCC TCAAATCACA AAAGAAAAAT TTAAACCTGG
P128-R              AGCCTTAGCA TCGGGTTATC ACAATCTTCC TCAAATCACA AAAGAAAAAT TTAAACCTGG
O213                AGCCTTAGCA TCGGGTTATC ACAATCTTCC TCAAATCACA AAAGAAAAAT TTAAACCTGG
O226                AGCCTTAGCA TCGGGTTATC ACAATCTTCC TCAAATCACA AAAGAAAAAT TTAAACCTGG
Cons-Plank98,7%     AGCCTTAGCA TCGGGTTATC ACAATCTTCC TCAAATCACA AAAGAAAAAT TTAAACCTGG
Nod-HEMclone        AGCATTAGCA TCAGGTTATC ACAACCTCCC CCAAATCACC GCAGAAAAAT TTCAACCTAG
Nod-BY1clone        AGCATTAGCA GCAGGTTATC ACAACCTCCC CCAAATCACC GCAGAAAAAT TTCAACCTAG
Nod-F81clone        AGCATTAGCA TCAGGTTATC ACAACCTCCC CCAAATCACC GCAGAAAAAT TTCAACCTAG
Cons-Nod99,4%       AGCATTAGCA TCAGGTTATC ACAACCTCCC CCAAATCACC GCAGAAAAAT TTCAACCTAG
Nostoc152           AGCATTAGCC GCAGGTTATC ATAATCTTCC CGACATCACT ACAGAAAAAT TTCAACCCAG

                    ....|....|  ....|....|  ....|....|  ....|....|  ....|....|  ....|....|
                        65          75          85          95         105         115
Ana90-Leo           CTTTTTAAAT TCAGAAAAAA TTCTCTTTAG AACGGGAGAT TTAGGTAAAC AAATTGCTCC
Ana299B             CTTTTTAAAT TCAGAAAAAA TTCTCTTTAG AACGGGAGAT TTAGGTAAAC AAATTGCTCC
Ana315              CTTTTTAAAT TCAGAAAAAA TTCTCTTTAG AACGGGAGAT TTAGGTAAAC AAATTGCTCC
AnaPH256            CTTTTTAAAT TCAGAAAAAA TTCTCTTTAG AACGGGAGAT TTAGGTAAAC AAATTGCTCC
Ana318              CTTTTTAAAT TCAGAAAAAA TTCTCTTTAG AACGGGAGAT TTAGGTAAAC AAATTGCTCC
Ana202A1            CTTTTTAAAT TCAGAAAAAA TTCTCTTTAG AACGGGAGAT TTAGGTAAAC AAATTGCTCC
Ana202A2-41         CTTTTTAAAT TCAGAAAAAA TTCTCTTTAG AACGGGAGAT TTAGGTAAAC AAATTGCTCC
Ana66A              CTTTTTAAAT TCAGGAAAAA TTCTCTTTAG AACGGGAGAT TTAGGTAAAC AAATTGCTCC
Ana83-1             CTTTTTAAAT TCAGAAAAAA TTCTCTTTAG AACGGGAGAT TTAGGTAAAC AAATTGCTCC
Cons-Ana98%         CTTTTTAAAT TCAGAAAAAA TTCTCTTTAG AACGGGAGAT TTAGGTAAAC AAATTGCTCC
Mpcc7806-Tillet     CTTTCTTGAT GAGACAAAAA CTCTCTTTAG AACCGGCGAT TTAGGCAAGC AAACTGCTCC
M205                CTTTCTTGAT GAGACAAAAA CTCTCTTTAG AACCGGCGAT TTAGGCAAAC AAACTGCTCC
M98                 CTTTCTTGAT GAGACAAAAA CTCTCTTTAG AACCGGCGAT TTAGGCAAGC AAACTGCTCC
Mgl260735           CTTTCTTGAT GAGACAAAAA CTCTCTTTAG AACCGGCGAT TTAGGCAAAC AAACTGCTCC
Mgl280646           CTTTCTTGAT GAGACAAAAA CTCTCTTTAG AACCGGCGAT TTAGGCAAGC AAACTGCTCC
Mpcc7941            CTTTCTTGAT GAGACAAAAA CTCTCTTTAG AACCGGCGAT TTAGGCAAGC AAACTGCTCC
Miz25               CTTTCTTGAT GAGACAAAAA CTCTCTTTAG AACCGGCGAT TTAGGCAAGC AAACTGCTCC
```

Fig. 19B

```
MizM5                CTTTCTTGAT GAGACAAAAA CTCTCTTTAG AACCGGCGAT TTAGGCAAGC AAACTGCTCC
Mnies102             CTTTCTTGAT GAGACAAAAA CTCTCTTTAG AACCGGCGAT TTAGGCAAGC AAACTGCTCC
MniesA89             CTTTCTTGAT GAGACAAAAA CTCTCTTTAG AACCGGCGAT TTAGGCAAGC AAACTGCTCC
Cons-Mic97,4%        CTTTCTTGAT GAGACAAAAA CTCTCTTTAG AACCGGCGAT TTAGGCAAGC AAACTGCTCC
P49                  CTTTTTTAAT CAGAAAACAA CGATGTTTAG AACCGGGGAT TTAGGGAAAC AAACTGCTCC
P97                  CTTTTTTGAT CAGAAAACAA CGATGTTTAG AACCGGGGAT TTAGGGAAAC AAACTGCTCC
P126-8               CTTTTTTGAT CAGAAAACAA CGATGTTTAG AACCGGGGAT TTAGGGAAAC AAACTGCTCC
P127                 CTTTTTTGAT CAGAAAACAA CGATGTTTAG AACCGGGGAT TTAGGGAAAC AAACTGCTCC
P128-R               CTTTTTTGAT CAGAAAACAA CGATGTTTAG AACCGGGGAT TTAGGGAAAC AAACTGCTCC
0213                 CTTTTTTAAT CAGAAAACAA CGATGTTTAG AACCGGGGAT TTAGGGAAAC AAACTGCTCC
0226                 CTTTTTTAAT CAGAAAACAA CGATGTTTAG AACCGGGGAT TTAGGGAAAC AAACTGCTCC
Cons-Plank98,7%      CTTTTTTGAT CAGAAAACAA CGATGTTTAG AACCGGGGAT TTAGGGAAAC AAACTGCTCC
Nod-HEMclone         CTTTATGACT GAGGGAAAAA CTATCTTTAG AACCGGAGAT TTAGGTAAAC AAATTGCCCC
Nod-BY1clone         CTTTATGACT GAGGGAAAAA CTATCTTTAG AACCGGAGAT TTAGGTAAAC AAATTGCCCC
Nod-F81clone         CTCTATGACT GAGGGAAAAA CTATCTTTAG AACCGGAGAT TTAGGTAAAC AAATTGCCCC
Cons-Nod99,4%        CTTTATGACT GAGGGAAAAA CTATCTTTAG AACCGGAGAT TTAGGTAAAC AAATTGCCCC
Nostoc152            CTTGATAAGT GAGGGAAAAA CTCTCTTTAG AACGGGAGAT TTAGGTAAAC AAACTGCTCC

                     ....|....| ....|....| ....|....| ....|....| ....|....| ....|....|
                          125        135        145        155        165        175
Ana90-Leo            AGGTGTGATT GAATTTATAG GGCGAAAAGA TAATCAAGTT AAGGTCAATG GCTATCGTGT
Ana299B              AGGTGTGATT GAATTTATAG GGCGAAAAGA TAATCAAGTT AAGGTCAATG GCTATCGTGT
Ana315               AGGTGTGATT GAATTTATAG GGCGAAAAGA TAATCAAGTT AAGGTCAATG GCTATCGTGT
AnaPH256             AGGTGTGATT GAATTTATAG GGCGAAAAGA TAATCAAGTT AAGGTCAATG GCTATCGTGT
Ana318               AGGTGTGATT GAATTTATAG GGCGAAAAGA TAATCAAGTT AAGGTCAATG GCTATCGTGT
Ana202A1             AGGTGTGATT GAATTTATAG GGCGAAAAGA TAATCAAGTT AAGGTCAATG GCTATCGTGT
Ana202A2-41          AGGTGTGATT GAATTTATAG GGCGAAAAGA TAATCAAGTT AAGGTCAATG GCTATCGTGT
Ana66A               AGGTGTGATT GAATTTATAG GGCGAAAAGA TAATCAAGTT AAGGTCAATG GCTATCGTGT
Ana83-1              AGGTGTGATT GAATTTATAG GGCGAAAAGA TAATCAAGTT AAGGTCAATG GCTATCGTGT
Cons-Ana98%          AGGTGTGATT GAATTTATAG GGCGAAAAGA TAATCAAGTT AAGGTCAATG GCTATCGTGT
Mpcc7806-Tillet      CGGTATCATT GAGTTTATGG GACGAAAAGA TAATCAAGTT AAGGTCAATG GTTATCGAAT
M205                 AGGTATCATT GAGTTTATGG GACGAAAAGA TAATCAAGTT AAGGTCAATG GTTATCGAAT
M98                  GGGTATCATT GAGTTTATGG GACGAAAAGA TAATCAAGTT AAGGTCAATG GTTATCGAAT
Mgl260735            AGGTATCATT GAGTTTATGG GACGAAAAGA TAATCAAGTT AAGGTCAATG GTTATCGAAT
Mgl280646            GGGTATCATT GAGTTTATGG GACGAAAAGA TAATCAAGTT AAGGTCAATG GTTATCGAAT
Mpcc7941             GGGTATCATT GAGTTTATGG GACGAAAAGA TAATCAAGTT AAGGTCAATG GTTATCGAAT
Miz25                CGGTATCATT GAGTTTATGG GACGAAAAGA TAATCAAGTT AAGGTCAATG GTTATCGAAT
MizM5                GGGTATCATT GAGTTTATGG GACGAAAAGA TAATCAAGTT AAGGTCAATG GTTATCGAAT
Mnies102             CGGTATCATT GAGTTTATGG GACGAAAAGA TAATCAAGTT AAGGTCAATG GTTATCGAAT
MniesA89             CGGTATCATT GAGTTTATGG GACGAAAAGA TAATCAAGTT AAGGTCAATG GTTATCGAAT
Cons-Mic97,4%        CGGTATCATT GAGTTTATGG GACGAAAAGA TAATCAAGTT AAGGTCAATG GTTATCGAAT
P49                  CGGTGTGATT GAATTTATGG GCAGAAAAGA CAATCAAGTT AAGGTAAATG GCTATCGTAT
P97                  CGGTGTAATT GAATTTATGG GCAGAAAAGA CAATCAAGTT AAGGTAAATG GCTATCGTAT
P126-8               CGGTGTGATT GAATTTATGG GCAGAAAAGA CAATCAAGTT AAGGTAAATG GCTATCGTAT
P127                 CGGTGTGATT GAATTTATGG GCAGAAAAGA CAATCAAGTT AAGGTAAATG GCTATCGTAT
P128-R               CGGTGTGATT GAATTTATGG GCAGAAAAGA CAATCAAGTT AAGGTAAATG GCTATCGTAT
0213                 CGGTGTGATT GAATTTATGG GCAGAAAAGA CAATCAAGTT AAGGTAAATG GCTATCGTAT
0226                 CGGTGTGATT GAATTTATGG GCAGAAAAGA CAATCAAGTT AAGGTAAATG GCTATCGTAT
Cons-Plank98,7%      CGGTGTGATT GAATTTATGG GCAGAAAAGA CAATCAAGTT AAGGTAAATG GCTATCGTAT
Nod-HEMclone         AGGCGTGATT GAATTTCTTG GTCGTAAAGA TAATCAAGTT AAGGTGAATG GCTATCGTAT
Nod-BY1clone         AGGCGTGATT GAATTTCTTG GTCGTAAAGA TAATCAAGTT AAGGTGAATG GCTATCGCAT
Nod-F81clone         AGGCGTGATT GAATTTCTTG GTCGTAAAGA TAATCAAGTT AAGGTGAATG GCTATCGTAT
Cons-Nod99,4%        AGGCGTGATT GAATTTCTTG GTCGTAAAGA TAATCAAGTT AAGGTGAATG GCTATCGTAT
Nostoc152            AGGTGTCATT GAATTTATTG GGCGTAAAGA TAATCAAGTA AAGGTGAATG GTTATCGTAT

                     ....|....| ....|....| ....|....| ....|....| ....|....| ....|....|
                          185        195        205        215        225        235
Ana90-Leo            AGATCCAGGA GAAATTGAAT ACCAAATTAG CCGCTATGCC GAGATTGAGA AAGCAATTGT
Ana299B              AGATCCAGGA GAAATTGAAT ACCAAATTAG CCGCTATGCC GAGATTGAGA AAGCAATTGT
Ana315               AGATCCAGGA GAAATTGAAT ACCAAATTAG CCGCTATGCC GAGATTGAGA AAGCAATTGT
AnaPH256             AGATCCAGGA GAAATTGAAT ACCAAATTAG CCGCTATGCC GAGATTGAGA AAGCAATTGT
```

Fig. 19C

```
Ana318              AGATCCAGGA GAAATTGAAT ACCAAATTAG CCGCTATGCC GAGATTGAGA AAGCAATTGT
Ana202A1            AGATCCAGGA GAACTTGAAT ACCAAATTAG CCGCTATGCC GAGATTGAGA AAGCAATTGT
Ana202A2-41         AGATCCAGGA GAACTTGAAT ACCAAATTAG CCGCTATGCC GAGATTGAGA AAGCAATTGT
Ana66A              AGATCCAGGA GAACTTGAAT ACCAAATTAG CCGCTATGCC GAGATTGAGA AAGCAATTGT
Ana83-1             AGATCCAGGA GAACTTGAAT ACCAAATTAG CCGCTATGCC GAGATTGAGA AAGCAATTGT
Cons-Ana98%         AGATCCAGGA GAAATTGAAT ACCAAATTAG CCGCTATGCC GAGATTGAGA AAGCAATTGT
Mpcc7806-Tillet     TGACCCCGGA GAAATTGAAT ATCAATTGAC TCGTTATGCT CCCATTGAAA GAGCGATTGT
M205                TGACCCCGGA GAAATTGAAT ATCAATTGAC TCGTTATGCT CCCATTGAAA GAGCGATTGT
M98                 TGACCCCGGA GAAATTGAAT ATCAATTGAC TCGTTATGCT CCCATTGAAA GAGCGATTGT
Mgl260735           TGACCCCGGA GAAATTGAAT ATCAATTGAC TCGTTATGCT CCCATTGAAA GAGCGATTGT
Mgl280646           TGACCCCGGA GAAATTGAAT ATCAATTGAC TCGTTATGCT CCCATTGAAA GAGCGATTGT
Mpcc7941            TGACCCCGGA GAAATTGAAT ATCAATTGAC TCGTTATGCT CCCATTGAAA GAGCGATTGT
Miz25               TGACCCCGGA GAAATTGAAT ATCAATTGAC TCGCTATGCT CCCATTGAAA GAGCGATTGT
MizM5               TGATCCCGGA GAAATTGAAT ATCAATTGAC TCGTTATGCT CCCATTGAAA GAGCGATTGT
Mnies102            TGACCCCGGA GAAATTGAAT ATCAATTGAC TCGTTATGCT CCCATTGAAA GAGCGATTGT
MniesA89            TGACCCCGGA GAAATTGAAT ATCAATTGAC TCGTTATGCT CCCATTGAAA GAGCGATTGT
Cons-Mic97,4%       TGACCCCGGA GAAATTGAAT ATCAATTGAC TCGTTATGCT CCCATTGAAA GAGCGATTGT
P49                 CGACCCCGAA GAAATTGAAT ATCAACTTAA TCGTTATCCT CAGATTGAGA GAGCTATTAT
P97                 CGACCCCGAA GAAATTGAAT ATCAACTTAA TCGTTATCCT CAGATTGAGA GAGCTATTAT
P126-8              CGACCCCGAA GAAATTGAAT ATCAACTTAA TCGTTATCCT CAGATTGAGA GAGCTATTAT
P127                CGACCCCGAA GAAATTGAAT ATCAACTTAA TCGTTATCCT CAGATTGAGA GAGCTATTAT
P128-R              CGACCCCGAA GAAATTGAAT ATCAACTTAA TCGTTATCCT CAGATTGAGA GAGCTATTAT
O213                CGACCCCGAA GAAATTGAAT ATCAACTTAA TCGTTATCCT CAGATTGAGA GAGCTATTAT
O226                CGACCCCGAA GAAATTGAAT ATCAACTTAA TCGTTATCCT CAGATTGAGA GAGCTATTAT
Cons-Plank98,7%     CGACCCCGAA GAAATTGAAT ATCAACTTAA TCGTTATCCT CAGATTGAGA GAGCTATTAT
Nod-HEMclone        AGATCCAGGA GAAATTGAAT ACCAACTCAG CCGCCATTCT CAAATTGAGA GAGCAATCGT
Nod-BY1clone        AGATCCAGGA GAAATTGAAT ACCAACTCAG CCGCCATTCT CAAATTGAGA GAGCAATCGT
Nod-F81clone        AGATCCAGGA GAAATTGAAT ACCAACTCAG CCGCCATTCT CAAATTGAGA GAGCAATCGT
Cons-Nod99,4%       AGATCCAGGA GAAATTGAAT ACCAACTCAG CCGCCATTCT CAAATTGAGA GAGCAATCGT
Nostoc152           AGATCCAGGA GAAATTGAAT ATCAACTCAG CCGTCATGCT CAGATTGAAA GAGCGATTAT

                    ....|....| ....|....| ....|....| ....|....| ....|....| ....|....|
                      245        255        265        275        285        295
Ana90-Leo           CTTACCTATA GAGGTAAATA ACCAAATTCA ATTATCTGCT TATTGTCAAA CTGATAAAGA
Ana299B             CTTACCTATA GAGGTAAATA ACCAAATTCA ATTATCTGCT TATTGTCAAA CTGATAAAGA
Ana315              CTTACCTATA GAGGTAAATA ACCAAATTCA ATTATCTGCT TATTGTCAAA CTGATAAAGA
AnaPH256            CTTACCTATA GAGGTAAATA ACCAAATTCA ATTATCTGCT TATTGTCAAA CTGATAAAGA
Ana318              CTTACCTATA GAGGTAAATA ACCAAATTCA ATTATCTGCT TATTGTCAAA CTGATAAAGA
Ana202A1            CTTACCTATA GAGGTAAATA ACCAAATTCA ATTATCTGCT TATTGTCAAA CTGATAAAGA
Ana202A2-41         CTTACCTATA GAGGTAAATA ACCAAATTCA ATTATCTGCT TATTGTCAAA CTGATAAAGA
Ana66A              CTTACCTATA GAGGTAAATA ACCAAATTCA ATTATCTGCT TATTGTCAAA CTGATAAAGA
Ana83-1             CTTACCTATA GAGGTAAATA ACCAAATTCA ATTATCTGCT TATTGTCAAA CTGATAAAGA
Cons-Ana98%         CTTACCTATA GAGGTAAATA ACCAAATTCA ATTATCTGCT TATTGTCAAA CTGATAAAGA
Mpcc7806-Tillet     TTTACCCGTT CAAGTTAATA ATCAAACTCA ATTATCTGCT TACTGTCAAA CAGACAAAAC
M205                TTTACCCGTT CAAGTTAATA ATCAAACTCA ATTATCTGCT TACTGTCAAA CAGACAAAAC
M98                 TTTACCCGTT CAAGTTAATA ATCAAACTCA ATTATCTGCT TACTGTCAAA CAGACAAAAC
Mgl260735           TTTACCCGTT CAAGTTAATA ATCAAACTCA ATTATCTGCT TACTGTCAAA CAGACAAAAC
Mgl280646           TTTACCCGTT CAAGTTAATA ATCAAACTCA ATTATCTGCT TACTGTCAAA CAGACAAAAC
Mpcc7941            TTTACCCGTT CAAGTTAATA ATCAAACTCA ATTATCTGCT TACTGTCAAA CAGACAAAAC
Miz25               TTTACCCGTT CAAGTTAATA ATCAAACTCA ATTATCTGCT TACTGTCAAA CAGACAAAAC
MizM5               TTTACCCATT CAAGTGAATA ATCAAACTCA ATTATCTGCT TACTGTCAAA CAGAAAAAAC
Mnies102            TTTACCCGTT CAAGTGAATA ATCAAACTCA ATTATCTGCT TACTGTCAAA CAGACAAAAC
MniesA89            TTTACCCGTT CAAGTGAATA ATCAAACTCA ATTATCTGCT TACTGTCAAA CAGACAAAAC
Cons-Mic97,4%       TTTACCCGTT CAAGTTAATA ATCAAACTCA ATTATCTGCT TACTGTCAAA CAGACAAAAC
P49                 TCTACCGATA TCAGTCAATA ATCAAACTCA ATTATCAGCC TATTGTCAAA CCGATAAACA
P97                 TCTACCGATA TCAGTCAATA ATCAAACTCA ATTATCAGCC TATTGTCAAA CAGATAAACA
P126-8              TCTACCGATA TCAGTCAATA ATCAAACTCA ATTATCAGCC TATTGTCAAA CAGATAAACA
P127                TCTACCGATA TCAGTCAATA ATCAAACTCA ATTATCAGCC TATTGTCAAA CAGATAAACA
P128-R              TCTACCGATA TCAGTCAATA ATCAAACTCA ATTATCAGCC TATTGTCAAA CAGATAAACA
O213                TCTACCGATA TCAGTCAATA ATCAAACTCA ATTATCAGCC TATTGTCAAA CCGATAAACA
O226                TCTACCGATA TCAGTCAATA ATCAAACTCA ATTATCAGCC TATTGTCAAA CCGATAAACA
```

Fig. 19D

```
Cons-Plank98,7%     TCTACCGATA TCAGTCAATA ATCAAACTCA ATTATCAGCC TATTGTCAAA CAGATAAACA
Nod-HEMclone        ATTGCCTACT AATGTAGATA ATCAAACCCA GTTATCAGCC TATTGTAAAA CTGAGTCAGA
Nod-BY1clone        ATTGCCTACT AATGTAGATA ATCAAACCCA GTTATCAGCC TATTGTAAAA CTGAGTCAGA
Nod-F81clone        ATTGCCTACT AATGTAGATA ATCAAACCCA GTTATCAGCC TATTGTAAAA CTGAGTCAGA
Cons-Nod99,4%       ATTGCCTACT AATGTAGATA ATCAAACCCA GTTATCAGCC TATTGTAAAA CTGAGTCAGA
Nostoc152           ATTGCCTATC AATGTAGATA ATCAAACTCA ATTATCTGCT TATTGTCAAA CTGATAAAGA

                    ....|....| ....|....| ....|....| ....|....| ....|....| ....|....|
                      305        315        325        335        345        355
Ana90-Leo           TATAAAATTT TCTGAAATCA GAGAATTTTT AGCTAAATAT TTGCCAGTTT ACATGATTCC
Ana299B             TATAAAATTT TCTGAAATCA GAGAATTTTT AGCTAAATAT TTGCCAGTTT ACATGATTCC
Ana315              TATAAAATTT TCTGAAATCA GAGAATTTTT AGCTAAATAT TTGCCAGTTT ACATGATTCC
AnaPH256            TATAAAATTT TCTGAAATCA GAGAATTTTT AGCTAAATAT TTGCCAGTTT ACATGATTCC
Ana318              TATAAAATTT TCTGAAATCA GAGAATTTTT AGCTAAATAT TTGCCAGTTT ACATGATTCC
Ana202A1            TATAAAAATT TCTGAAATCA GAGAATTTTT AGCTAAATAT TCGCCAGTTT ACATGATTCC
Ana202A2-41         TATAAAAATT TCTGAAATCA GAGAATTTTT AGCTAAATAT TCGCCAGTTT ACATGATTCC
Ana66A              TATAAAAATT TCTGAAATCA GAGAATTTTT AGCTAAATAT TCGCCAGTTT ACATGATTCC
Ana83-1             TATAAAAATT TCTGAAATCA GAGAATTTTT AGCTAAATAT TCGCCAGTTT ACATGATTCC
Cons-Ana98%         TATAAAATTT TCTGAAATCA GAGAATTTTT AGCTAAATAT TTGCCAGTTT ACATGATTCC
Mpcc7806-Tillet     TCTAGAAATT GCTGAGATTC GAGAATTACT TGCCAAATTT TTACCAGTTT ATATGATTCC
M205                TCTAGAAATT GCTGAGATTC GAGAATTACT TGCCAAATTT TTACCAGTTT ATATGATTCC
M98                 TCTAGAAATT GCTGAGATTC GAGAATTACT TGCCAAATTT TTACCAGTTT ATATGATTCC
Mgl260735           TCTAGAAATT GCTGAGATTC GAGAATTACT TGCCAAATTT TTACCAGTTT ATATGATTCC
Mgl280646           TCTAGAAATT GCTGAGATTC GAGAATTACT TGCCAAATTT TTACCAGTTT ATATGATTCC
Mpcc7941            TCTAGAAATT GCTGAGATTC GAGAATTACT TGCCAAATTT TTACCAGTTT ATATGATTCC
Miz25               TCTAGAAATT GCTGAGATTC GAGAATTACT TGCCAAATTT TTACCAGTTT ATATGATTCC
MizM5               TCTAGAAATT GCTGAGATTC GAGAATTTCT TGCCAAGTTT TTACCAGTTT ATATGATTCC
Mnies102            TCTAGAAATT GCTGAGATTC GAGAATTTCT TGCCAAGTTT TTGCCAGTTT ATATGATTCC
MniesA89            TCTAGAAATT GCTGAGATTC GAGAATTTCT TGCCAAGTTT TTGCCAGTTT ATATGATTCC
Cons-Mic97,4%       TCTAGAAATT GCTGAGATTC GAGAATTACT TGCCAAATTT TTACCAGTTT ATATGATTCC
P49                 GATAGAAATT TCTGAAATCA GAGAATTTCT AGCTAATTTT CTGCCAGTTT ACATGATTCC
P97                 GATAGAAATT TCTGAAATTA GAGAATTTCT AGCTAATTTT CTGCCAGTTT ACATGATTCC
P126-8              GATAGAAATT TCTGAAATTA GAGAATTTCT AGCTAATTTT TTGCCAGTTT ACATGATTCC
P127                GATAGAAATT TCTGAAATTA GAGAATTTCT AGCTAATTTT TTGCCAGTTT ACATGATTCC
P128-R              GATAGAAATT TCCGAAATTA GAGAATTTTT AGCTAATTTT CTGCCAGTTT ACATGATTCC
0213                GATAGAAATT TCTGAAATCA GAGAATTTCT AGCTAATTTT CTGCCAGTTT ACATGATTCC
0226                GATAGAAATT TCTGAAATCA GAGAATTTCT AGCTAATTTT CTGCCAGTTT ACATGATTCC
Cons-Plank98,7%     GATAGAAATT TCTGAAATTA GAGAATTTCT AGCTAATTTT CTGCCAGTTT ACATGATTCC
Nod-HEMclone        CATAGAAATT TCCGAAATTC GAGAATTTCT ATCGAACTTT TTGCCCGTTT ACATGATTCC
Nod-BY1clone        CATAGAAATT TCCGAAATTC GAGAATTTCT ATCGAACTTT TTGCCCGTTT ACATGATTCC
Nod-F81clone        CATAGAAATT TCCGAAATTC GAGAATTTCT ATCGAACTTT TTGCCCGTTT ATATGATTCC
Cons-Nod99,4%       CATAGAAATT TCCGAAATTC GAGAATTTCT ATCGAACTTT TTGCCCGTTT ACATGATTCC
Nostoc152           CATCGAAATT GCTGAAATTA GAGAATTTCT CTCTAAATTT TTGCCAGTTT ATATGATTCC

                    ....|....| ....|....| ....|....| ....|....| ....|....| ....|....|
                      365        375        385        395        405        415
Ana90-Leo           TAGTTCCTTT ATCTTCTTAA AGCAATTTCC CTTAACTAAA CATGGCAAAT TTGACTTGCG
Ana299B             TAGTTCCTTT ATCTTCTTAA AGCAATTTCC CTTAACTAAA CATGGCAAAT TTGACTTGCG
Ana315              TAGTTCCTTT ATCTTCTTAA AGCAATTTCC CTTAACTAAA CATGGCAAAT TTGACTTCCG
AnaPH256            TAGTTCCTTT ATCTTCTTAA AGCAATTTCC CTTAACTAAA CATGGCAAAT TTGACTTCCG
Ana318              TAGTTCCTTT ATCTTCTTAA AGCAATTTCC CTTAACTAAA CATGGCAAAT TTGACTTGCG
Ana202A1            TAGTTCCTTT ATCTTCTTAA AGCAATTTCC CTTAACTAAA CATGGCAAAC TTGACTTGCG
Ana202A2-41         TAGTTCCTTT ATCTTCTTAA AGCAATTTCC CTTAACTAAA CATGGCAAAC TTGACTTGCG
Ana66A              TAGTTCCTTT ATCTTCTTAA AGCAATTTCC CTTAACTAAA CATGGCAAAC TTGACTTGCG
Ana83-1             TAGTTCCTTT ATCTTCTTAA AGCAATTTCC CTTAACTAAA CATGGCAAAC TTGACTTGCG
Cons-Ana98%         TAGTTCCTTT ATCTTCTTAA AGCAATTTCC CTTAACTAAA CATGGCAAAT TTGACTTGCG
Mpcc7806-Tillet     GAGTTACTTT ATTTTTTTAA AGCAATTCCC CTTAACTCGA CATGGGAAAC TTGACCTGCA
M205                GAGTTACTTT ATTTTTTTAA AGCAATTCCC CTTAACTCGA CATGGAAAAC TTGACCTGCA
M98                 GAGTTACTTT ATTTTTTTAA AGCAATTCCC CTTAACTCGA CATGGAAAAC TTGACCTGCA
Mgl260735           GAGTTACTTT ATTTTTTTAA AGCAATTCCC CTTAACTCGA CATGGAAAAC TTGACCTGCA
Mgl280646           GAGTTACTTT ATTTTTTTAA AGCAATTCCC CTTAACTCGA CATGGAAAAC TTGACCTGCA
```

Fig. 19E

```
Mpcc7941              GAGTTACTTT ATTTTTTTAA AGCAATTCCC CTTAACTCGA CATGGAAAAC TTGACCTGCA
Miz25                 GAGTTACTTT ATTTTTTTAA AGCAATTCCC CTTAACTCGA CATGGGAAAC TTGACCTGCA
MizM5                 CAGTTACTTT ATTTTTTTAA AGCAATTCCC TTTAACTCGA CATGGGAAAC TTGACCTGCA
Mnies102              CAGTTACTTT ATTTTTTTAA AGCAATTCCC CTTAACTCGA CATGGGAAAC TTGACCTGCA
MniesA89              CAGTTACTTT ATTTTTTTAA AGCAATTCCC CTTAACTCGA CATGGGAAAC TTGACCTGCA
Cons-Mic97,4%         GAGTTACTTT ATTTTTTTAA AGCAATTCCC CTTAACTCGA CATGGAAAAC TTGACCTGCA
P49                   TAGTTACTTT ATTTTCTTAA AGCAATTCCC CCTAACTAAA CACGGCAAAC TTGACTTAAA
P97                   TAGTTACTTT ATTTTCTTAA AGCAATTTCC CCTAACTAAA CACGGCAAAC TTGACTTAAA
P126-8                TAGTTACTTT ATTTTCTTAA AGCAATTTCC CCTAACTAAA CACGGCAAAC TTGACTTAAA
P127                  TAGTTACTTT ATTTTCTTAA AGCAATTTCC CCTAACTAAA CACGGCAAAC TTGACTTAAA
P128-R                TAGTTACTTT ATTTTCTTAA AGCAATTTCC CCTAACTAAA CACGGCAAAC TTGACTTAAA
O213                  TAGTTACTTT ATTTTCTTAA AGCAATTCCC CCTAACTAAA CACGGCAAAC TTGACTTAAA
O226                  TAGTTACTTT ATTTTCTTAA AGCAATTCCC CCTAACTAAA CACGGCAAAC TTGACTTAAA
Cons-Plank98,7%       TAGTTACTTT ATTTTCTTAA AGCAATTTCC CCTAACTAAA CACGGCAAAC TTGACTTAAA
Nod-HEMclone          TACTTTCTTT ATCTTCTTAA AGCAATTTCC CTTAACCAGA CATGGGAAAA TTGATTTGCG
Nod-BY1clone          TACTTTCTTT ATCTTCTTAA AGCAATTTCC CTTAACCAGA CATGGGAAAA TTGATTTGCG
Nod-F81clone          TACTTTCTTT ATCTTCTTAA AGCAATTTCC CTTAACCAGA CATGGGAAAA TTGATTTGCG
Cons-Nod99,4%         TACTTTCTTT ATCTTCTTAA AGCAATTTCC CTTAACCAGA CATGGGAAAA TTGATTTGCG
Nostoc152             TACTTCCTTT ATCTTCCTCA AGCAATTTCC TCTAACCAGA CATGGCAAAA TTGACTTGCG

                      ....|....| ....|....| ....|....| ....|....| ....|....| ....|....|
                           425        435        445        455        465        475
Ana90-Leo             ATCGCTCATC GCTCTCAAGC CAACAGATCA ATTAACACAA GTCTC---TT ATACTGCACC
Ana299B               ATCGCTCATC GCTCTCAAGC CAACAGATCA ATTAACACAA GTCTC---TT ATACTGCACC
Ana315                ATCGCTCATC GCTCTCAAGC CAACAGATCA ATTAACACAA GTCTC---TT ATACTGCACC
AnaPH256              ATCGCTCATC GCTCTCAAGC CAACAGATCA ATTAACACAA GTCTC---TT ATACTGCACC
Ana318                ATCGCTCATC GCTCTCAAGC CAACAGATCA ATTAACACAA GTCTC---TT ATACTGCACC
Ana202A1              ATCGCTCGTC GCTCTCAAGC CGACAGATCA ACTAACACAA GTCTC---TT ATACTGCACC
Ana202A2-41           ATCGCTCGTC GCTCTCAAGC CGACAGATCA ACTAACACAA GTCTC---TT ATACTGCACC
Ana66A                ATCGCTCGTC GCTCTCAAGC CGACAGATCA ACTAACACAA GTCTC---TT ATACTGCACC
Ana83-1               ATCGCTCGTC GCTCTCAAGC CGACAGATCA ACTAACACAA GTCTC---TT ATACTGCACC
Cons-Ana98%           ATCGCTCATC GCTCTCAAGC CAACAGATCA ATTAACACAA GTCTC---TT ATACTGCACC
Mpcc7806-Tillet       CTCCCTGAGA GAACTCAGAG AAACTGGTAA ATCTCTGGTG AATTCTAATT ATGTTGCACC
M205                  CTCCCTGAGA GAACTCAGAG AAACTGGTAA ATCTCTGGTG AATTCTAATT ACGTTGCACC
M98                   CTCCCTGAGA GAACTCAGAG AAACTGGTAA ATCTCTGGTG AATTCTAATT ACGTTGCACC
Mgl260735             CTCCCTGAGA GAACTCAGAG AAACTGGTAA ATCTCTGGTG AATTCTAATT ACGTTGCACC
Mgl280646             CTCCCTGAGA GAACTCAGAG AAACTGGTAA ATCTCTGGTG AATTCTAATT ACGTTGCACC
Mpcc7941              CTCCCTGAGA GAACTCAGAG AAACTGGTAA ATCTCTGGTG AATTCTAATT ACGTTGCACC
Miz25                 CTCCCTGAGA GAACTCAGAG AAACTGGTAA ATCTCTGGTG AATTCTAATT ATGTTGCACC
MizM5                 CTCCCTGAGA CAACTCAGAG AAACCGGTAA ATATCTGGTT AATTCTAATT ATGTTGCACC
Mnies102              CTCCCTGAGA GAACTCAAAG AAACCAGTAA ATCTCTGGTT AATTCTAATT ATGTTGCACC
MniesA89              CTCCCTGAGA GAACTCAAAG AAACCAGTAA ATCTCTGGTT AATTCTAATT ATGTTGCACC
Cons-Mic97,4%         CTCCCTGAGA GAACTCAGAG AAACTGGTAA ATCTCTGGTG AATTCTAATT ACGTTGCACC
P49                   CTCACTGATT GCACTCAATG AAACCGGGAA ATCTACCCAG GTA---AATT ATGTTGCACC
P97                   CTCAATGATT GCACTCAATG AAACCGGGAA ATCTACCCAA GTA---AATT ATGTTGCACC
P126-8                CTCAATGATT GCACTCAATG AAACCGGGAA ATCTACCCAA GTA---AATT ATGTTGCACC
P127                  CTCAATGATT GCACTCAATG AAACCGGGAA ATCTACCCAA GTA---AATT ATGTTGCACC
P128-R                CTCAATGATT GCACTCAATG AAACCGGGAA ATCTACCCAA GTA---AATT ATGTTGCACC
O213                  CTCACTGATT GCACTCAATG AAACCGGGAA ATCTACCCAG GTA---AATT ATGTTGCACC
O226                  CTCACTGATT GCACTCAATG AAACCGGGAA ATCTACCCAG GTA---AATT ATGTTGCACC
Cons-Plank98,7%       CTCAATGATT GCACTCAATG AAACCGGGAA ATCTACCCAA GTA---AATT ATGTTGCACC
Nod-HEMclone          ATCCCTGGCT GAATTCAAGG GAATAGGTAA CTTAACACAG TTAGC---GT ATACTGCACC
Nod-BY1clone          ATCCCTGGCT GAATTCAAGG GAATAGGTAA CTTAACACAG TTAGC---GT ATACTGCACC
Nod-F81clone          ATCCCTGGCT GAATTCAAGG GAATAGGTAA CTTAACACAG TTAGC---GT ATACTGCACC
Cons-Nod99,4%         ATCCCTGGCT GAATTCAAGG GAATAGGTAA CTTAACACAG TTAGC---GT ATACTGCACC
Nostoc152             ATCGCTTGCT GAACTCCAAG GAATCGGTAA GTTAACACAG GCAGA---TT ATACTGCACC

                      ....|....| ....|....| ....|....| ....|....| ....|....| ....|....|
                           485        495        505        515        525        535
Ana90-Leo             GCGTAATACT TTAGAATCAA AGCTAGTCCA TATTTGGGAA AAAATTCTCA CTAAACATCC
Ana299B               GCGTAATACT TTAGAATCAA AGCTAGTCCA TATTTGGGAA AAAATTCTCA CTAAACATCC
```

Fig. 19F

```
Ana315              GCGTAATACT TTAGAATCAA AGCTAGTCCA TATTTGGGAA AAAATTCTCA CTAAACATCC
AnaPH256            GCGTAATACT TTAGAATCAA AGCTAGTCCA TATTTGGGAA AAAATTCTCA CTAAACATCC
Ana318              GCGTAATACT TTAGAATCAA AGCTAGTCCA TATTTGGGAA AAAATTCTCA CTAAACATCC
Ana202A1            GCGTAATACT TTAGAATCAA AGCTAGTCCA TATTTGGGAA AAAATTCTCA CTAAACATCC
Ana202A2-41         GCGTAATACT TTAGAATCAA AGCTAGTCCA TATTTGGGAA AAAATTCTCA CTAAACATCC
Ana66A              GCGTAATACT TTAGAATCAA AGCTAGTCCA TATTTGGGAA AAAATTCTCA CTAAACATCC
Ana83-1             GCGTAATACT TTAGAATCAA AGCTAGTCCA TATTTGGGAA AAAATTCTCA CTAAACATCC
Cons-Ana98%         GCGTAATACT TTAGAATCAA AGCTAGTCCA TATTTGGGAA AAAATTCTCA CTAAACATCC
Mpcc7806-Tillet     CCGGAATTAT TTAGAATCCA ATCTCGTTAG TATCTGGGAA AAAATTCTCT CTAAACATCC
M205                CCGGAATTAT TTAGAATCCA ATCTCGTTAG TATCTGGGAA AAAATTCTCT CTAAACATCC
M98                 CCGGAATTAT TTAGAATCCA ATCTCGTTAG TATCTGGGAA AAAATTCTCT CTAAACATCC
Mgl260735           CCGGAATTAT TTAGAATCCA ATCTCGTTAG TATCTGGGAA AAAATTCTCT CTAAACATCC
Mgl280646           CCGGAATTAT TTAGAATCCA ATCTCGTTAG TATCTGGGAA AAAATTCTCT CTAAACATCC
Mpcc7941            CCGGAATTAT TTAGAATCCA ATCTCGTTAG TATCTGGGAA AAAATTCTCT CTAAACATCC
Miz25               CCGGAATTAT TTAGAATCCA ATCTCGTTAG TATCTGGGAA AAAATTCTCT CTAAACATCC
MizM5               CCGTAATCAT TTAGAATCCA ATCTCGTTAG TATCTGGGAA AAAATTCTCT CTAAACATCC
Mnies102            CCGTAATCAT TTAGAATCCA ATCTCGTTAG TATCTGGGAA AAAATTCTCT CTAAACATCC
MniesA89            CCGTAATCAT TTAGAATCCA ATCTCGTTAG TATCTGGGAA AAAATTCTCT CTAAACATCC
Cons-Mic97,4%       CCGGAATTAT TTAGAATCCA ATCTCGTTAG TATCTGGGAA AAAATTCTCT CTAAACATCC
P49                 GCGTAATAAT TTAGAGTCAA ACCTAGTTAG AATCTGGGAA AAGATTCTGA CCAAACATCC
P97                 GCGTAATAAT TTAGAGTCAA ACCTAGTTAG AATCTGGGAA AAGATTCTGA CCAAACATCC
P126-8              GCGTAATAAT TTAGAGTCAA ACCTAGTTAG AATCTGGGAA AAGATTCTAA CCAAACATCC
P127                GCGTAATAAT TTAGAGTCAA ACCTAGTTAG AATCTGGGAA AAGATTCTAA CCAAACATCC
P128-R              GCGTAATAAT TTAGAGTCAA ACCTAGTTAG AATCTGGGAA AAGATTCTAA CCAAACATCC
O213                GCGTAATAAT TTAGAGTCAA ACCTAGTTAG AATCTGGGAA AAGATTCTGA CCAAACATCC
O226                GCGTAATAAT TTAGAGTCAA ACCTAGTTAG AATCTGGGAA AAGATTCTGA CCAAACATCC
Cons-Plank98,7%     GCGTAATAAT TTAGAGTCAA ACCTAGTTAG AATCTGGGAA AAGATTCTGA CCAAACATCC
Nod-HEMclone        GCGCAATAAT TTAGAGTCCA AGCTCGTACA TATTTGGGAA AAAATTCTCA CCAAACAACC
Nod-BY1clone        GCGCAATAAT TTAGAGTCCA AGCTCGTACA TATTTGGGAA AAAATTCTCA CCAAACAACC
Nod-F81clone        GCGCAATAAT TTAGAGTCCA AGCTCGTACA TATTTGGGAA AAAATTCTCA CCAAACAACC
Cons-Nod99,4%       GCGCAATAAT TTAGAGTCCA AGCTCGTACA TATTTGGGAA AAAATTCTCA CCAAACAACC
Nostoc152           GCGCAATGAT TTAGAGTCCA AGCTAGTAAA GATTTGGGAA AAAATTCTCA CCACACATCC

                    ....|....| ....|....| ....|....| ....|....| ....|....| ....|....|
                          545        555        565        575        585        595
Ana90-Leo           CATTGGAATT TTTGATAACT TTTTTGAAAT CGGCGGACAC TCTCTGCTCC TTTCTAGAGT
Ana299B             CATTGGAATT TTTGATAACT TTTTTGAAAT CGGCGGACAC TCTCTGCTCC TTTCTAGAGT
Ana315              CATTGGAATT TTTGATAACT TTTTTGAAAT CGGCGGACAC TCTCTGCTCC TTTCTAGAGT
AnaPH256            CATTGGAATT TTTGATAACT TTTTTGAAAT CGGCGGACAC TCTCTGCTCC TTTCTAGAGT
Ana318              CATTGGAATT TTTGATAACT TTTTTGAAAT CGGCGGACAC TCTCTGCTCC TTTCTAGAGT
Ana202A1            CATTGGAATT TTTGATAACT TTTTTGAAAT CGGCGGACAC TCTCTGCTCC TTTCTAGAGT
Ana202A2-41         CATTGGAATT TTTGATAACT TTTTTGAAAT CGGCGGACAC TCTCTGCTCC TTTCTAGAGT
Ana66A              CATTGGAATT TTTGATAACT TTTTTGAAAT CGGCGGACAC TCTCTGCTCC TTTCTAGAGT
Ana83-1             CATTGGAATT TTTGATAACT TTTTTGAAAT CGGCGGACAC TCTCTGCTCC TTTCTAGAGT
Cons-Ana98%         CATTGGAATT TTTGATAACT TTTTTGAAAT CGGCGGACAC TCTCTGCTCC TTTCTAGAGT
Mpcc7806-Tillet     TATCGGTATT TTTGATAACT TCTTTGAAAT TGGCGGTCAT TCTCTACTCT TATCAAGGGT
M205                TATCGGTATT TTTGATAACT TCTTTGAAAT TGGCGGTCAT TCTCTACTCT TATCAAGGGT
M98                 TATCGGTATT TTTGATAACT TCTTTGAAAT TGGCGGTCAT TCTCTACTCT TATCAAGGGT
Mgl260735           TATCGGTATT TTTGATAACT TCTTTGAAAT TGGCGGTCAT TCTCTACTCT TATCAAGGGT
Mgl280646           TATCGGTATT TTTGATAACT TCTTTGAAAT TGGCGGTCAT TCTCTACTCT TATCAAGGGT
Mpcc7941            TATCGGTATT TTTGATAACT TCTTTGAAAT TGGCGGTCAT TCTCTACTCT TATCAAGGGT
Miz25               TATCGGTATT TTTGATAACT TCTTTGAAAT TGGCGGTCAT TCTCTACTCT TATCAAGGGT
MizM5               TATCGGTATT TTTGATAACT TCTTTGAAAT TGGCGGTCAT TCTCTACTCT TATCAAGGGT
Mnies102            TATCGGTATT TTTGATAACT TTTTTGAAAT TGGTGGTCAT TCTCTACTCT TATCAAGGGT
MniesA89            TATCGGTATT TTTGATAACT TCTTTGAAAT TGGCGGTCAT TCTCTACTCT TATCAAGGGT
Cons-Mic97,4%       TATCGGTATT TTTGATAACT TCTTTGAAAT TGGCGGTCAT TCTCTACTCT TATCAAGGGT
P49                 CATCGGTATT TTTGATAACT TCTTTGAAAT TGGCGGACAT TCTCTGATGC TTTCGAGAAT
P97                 CATCGGTATT TTTGATAACT TCTTTGAAAT TGGCGGACAT TCTCTGATGC TTTCGAGAAT
P126-8              CATCGGTATT TTTGATAACT TCTTTGAAAT TGGCGGACAT TCTCTGATGC TTTCGAGAAT
P127                CATCGGTATT TTTGATAACT TCTTTGAAAT TGGCGGACAT TCTCTGATGC TTTCGAGAAT
P128-R              CATCGGTATT TTTGATAACT TCTTTGAAAT TGGCGGACAT TCTCTGATGC TTTCGAGAAT
```

Fig. 19G

```
O213                CATCGGTATT TTTGATAACT TCTTTGAAAT TGGCGGACAT TCTCTGATGC TTTCGAGAAT
O226                CATCGGTATT TTTGATAACT TCTTTGAAAT TGGCGGACAT TCTCTGATGC TTTCGAGAAT
Cons-Plank98,7%     CATCGGTATT TTTGATAACT TCTTTGAAAT TGGCGGACAT TCTCTGATGC TTTCGAGAAT
Nod-HEMclone        CATTGGCATT TTTGATAACT TCTTTGAAAT TGGTGGACAC TCACTGCTGC TTTCCAGAGT
Nod-BY1clone        CATTGGCATT TTTGATAACT TCTTTGAAAT TGGTGGACAC TCACTGCTGC TTTCCAGAGT
Nod-F81clone        CATTGGCATT TTTGATAACT TCTTTGAAAT TGGTGGACAC TCACTGCTGC TTTCCAGAGT
Cons-Nod99,4%       CATTGGCATT TTTGATAACT TCTTTGAAAT TGGTGGACAC TCACTGCTGC TTTCCAGAGT
Nostoc152           CATCGGCATT TTTGATAACT TCTTTGAAAT TGGTGGACAC TCGCTGCTGC TTTCGAGAGT

                    ....|....| ....|....| ....|....| ....|....| ....|....| ....|....|
                          605        615        625        635        645        655
Ana90-Leo           AGTCACTCAC GTCCATAAAG AATTAAATGT ATTAGTTAAA TTAGCTGATT TCTTCAAAGT
Ana299B             AGTCACTCAC GTCCATAAAG AATTAAATGT ATTAGTTAAA TTAGCTGATT TCTTCAAAGT
Ana315              AGTCACTCAC GTCCATAAAG AATTAAATGT ATTAGTTAAA TTAGCTGATT TCTTCAAAGT
AnaPH256            AGTCACTCAC GTCCATAAAG AATTAAATGT ATTAGTTAAA TTAGCTGATT TCTTCAAAGT
Ana318              AGTCACTCAC GTCCATAAAG AATTAAATGT ATTAGTTAAA TTAGCTGATT TCTTCAAAGT
Ana202A1            AGTCACTCAC GTCCATAAAG AATTAAATGT ATTAGTTAAA TTAGCTGATT TCTTCAAAGT
Ana202A2-41         AGTCACTCAC GTCCATAAAG AATTAAATGT ATTAGTTAAA TTAGCTGATT TCTTCAAAGT
Ana66A              AGTCACTCAC GTCCATAAAG AATTAAATGT ATTAGTTAAA TTAGCTGATT TCTTCAAAGT
Ana83-1             AGTCACTCAC GTCCATAAAG AATTAAATGT ATTAGTTAAA TTAGCTGATT TCTTCAAAGT
Cons-Ana98%         AGTCACTCAC GTCCATAAAG AATTAAATGT ATTAGTTAAA TTAGCTGATT TCTTCAAAGT
Mpcc7806-Tillet     TGTAACCCGG GTTCATAAAG AACTAAATGT ATCCGTAAAA TTGGCTGACT TCTTTAAAGT
M205                TGTAACCCGG GTTCATAAAG AACTAAATGT ATCCGTAAAA TTAGCTGACT TCTTTAAAGT
M98                 TGTAACCCGG GTTCATAAAG AACTAAATGT ATCCGTAAAA TTAGCTGACT TCTTTAAAGT
Mgl260735           TGTAACCCGG GTTCATAAAG AACTAAATGT ATCCGTAAAA TTAGCTGACT TCTTTAAAGT
Mgl280646           TGTAACCCGG GTTCATAAAG AACTAAATGT ATCCGTAAAA TTAGCTGACT TCTTTAAAGT
Mpcc7941            TGTAACCCGG GTTCATAAAG AACTAAATGT ATCCGTAAAA TTAGCTGACT TCTTTAAAGT
Miz25               TGTAACCCGG GTTCATAAAG AACTAAATGT ATCCGTAAAA TTGGCTGACT TCTTTAAAGT
MizM5               TGTAACCCGG GTTCATAAAG AACTAAATGT ATCCGTAAAA TTGGCTGACT TCTTTAAAGT
Mnies102            TGTAACCCGG GTTCATAAAG AACTAAATGT ATCCGTAAAA TTAGCTGACT TCTTTAAAGT
MniesA89            TGTAACCCGG GTTCATAAAG AACTAAATGT ATCCGTAAAA TTGGCTGACT TCTTTAAAGT
Cons-Mic97,4%       TGTAACCCGG GTTCATAAAG AACTAAATGT ATCCGTAAAA TTAGCTGACT TCTTTAAAGT
P49                 CGTAACCCAC GTTCATAAAG AATTAAATGT ATCGGTAAAA TTGGCTGACT TCTTTAAAGT
P97                 CGTAACCCAC GTTCATAAAG AATTAAATGT ATCGGTAAAA TTGGCTGACT TCTTTAAAGT
P126-8              CGTAACCCAC GTTCATAAAG AATTAAATGT ATCGGTAAAA TTGGCTGACT TCTTTAAAGT
P127                CGTAACCCAC GTTCATAAAG AATTAAATGT ATCGGTAAAA TTGGCTGACT TCTTTAAAGT
P128-R              CGTAACCCAC GTTCATAAAG AATTAAATGT ATCGGTAAAA TTGGCTGACT TCTTTAAAGT
O213                CGTAACCCAC GTTCATAAAG AATTAAATGT ATCGGTAAAA TTGGCTGACT TCTTTAAAGT
O226                CGTAACCCAC GTTCATAAAG AATTAAATGT ATCGGTAAAA TTGGCTGACT TCTTTAAAGT
Cons-Plank98,7%     CGTAACCCAC GTTCATAAAG AATTAAATGT ATCGGTAAAA TTGGCTGACT TCTTTAAAGT
Nod-HEMclone        GGTAACTCAC GTTCATAAAG AATTAAATGT GTTGGTAAAA TTGGCTGAAT TCTTTAAAGT
Nod-BY1clone        GGTAACTCAC GTTCATAAAG AATTAAATGT GTTGGTAAAA TTGGCTGAAT TCTTTAAAGT
Nod-F81clone        GGTAACTCAC GTTCATAAAG AATTAAATGT GTTGGTAAAA TTGGCTGAAT TCTTTAAAGT
Cons-Nod99,4%       GGTAACTCAC GTTCATAAAG AATTAAATGT GTTGGTAAAA TTGGCTGAAT TCTTTAAAGT
Nostoc152           GGTAACTTAC GTTCATAAAG AATTAAATGT GTTAGTCAAA TTGGCTGACT TTTTTAAAGT

                    ....|....| ....|....| ....|....| ....|....| ....|....| ....|....|
                          665        675        685        695        705        715
Ana90-Leo           TCCCACAATT CTTGGATTAG CAGCTTTAAT ATCTAAAGCT CAATCTAACT ATCAAGAACC
Ana299B             TCCCACAATT CTTGGATTAG CAGCTTTAAT ATCTAAAGCT CAATCTAACT ATCAAGAACC
Ana315              TCCCACAATT CTTGGATTAG CAGCTTTAAT ATCTAAAGCT CAATCTAACT ATCAAGAACC
AnaPH256            TCCCACAATT CTTGGATTAG CAGCTTTAAT ATCTAAAGCT CAATCTAACT ATCAAGAACC
Ana318              TCCCACAATT CTTGGATTAG CAGCTTTAAT ATCTAAAGCT CAATCTAACT ATCAAGAACC
Ana202A1            TCCCACAATT CTTGGATTAG CAGCTTTAAT ATCTAAAGCT CAATCTAACT ATCAAGAACC
Ana202A2-41         TCCCACAATT CTTGGATTAG CAGCTTTAAT ATCTAAAGCT CAATCTAACT ATCAAGAACC
Ana66A              TCCCACAATT GTTGGATTAG CAGCTTTAAT ATCTAAAGCT CAATCTAACT ATCAAGAACC
Ana83-1             TCCCACAATT CTTGGATTAG CAGCTTTAAT ATCTAAAGCT CAATCTAACT ATCAAGAACC
Cons-Ana98%         TCCCACAATT CTTGGATTAG CAGCTTTAAT ATCTAAAGCT CAATCTAACT ATCAAGAACC
Mpcc7806-Tillet     TCCAACCATT GCTGGATTGG CTACTTTAAT CTCCCAGACT CAATACAATT ATCAAGAACC
M205                TCCAACCGTT GCTGGATTGG CGACTTTAAT CTCCCAGACT CAATACAATT ATCAAGAACC
M98                 TCCAACCGTT GCTGGATTGG CGACTTTAAT CTCCCAGACT CAATACAATT ATCAAGAACC
```

Fig. 19H

```
Mgl260735          TCCAACCGTT GCTGGATTGG CGACTTTAAT CTCCCAGACT CAATACAATT ATCAAGAACC
Mgl280646          TCCAACCGTT GCTGGATTGG CGACTTTAAT CTCCCAGACT CAATACAATT ATCAAGAACC
Mpcc7941           TCCAACCGTT GCTGGATTGG CGACTTTAAT CTCCCAGACT CAATACAATT ATCAAGAACC
Miz25              TCCAACCATT GCTGGATTGG CGACTTTAAT CTCCCAGACT CAATACAATT ATCAAGAACC
MizM5              TCCAACCATT GCTGGATTGG CGACTTTAAT CTCCCAGACT CAATACAATT ATCAAGAACC
Mnies102           TCCAACCGTT GCTGGATTGG CGACTTTAAT CTCCCAGACT CAATACAATT ATCAAGAACC
MniesA89           TCCAACCATT GCTGGATTGG CGACTTTAAT CTCCCAGACT CAATACAATT ATCAAGAACC
Cons-Mic97,4%      TCCAACCGTT GCTGGATTGG CGACTTTAAT CTCCCAGACT CAATACAATT ATCAAGAACC
P49                TCCTACCATT GCCGGATTAG CCGTTTTAGT CTCTAAAACT GAATATGATT ATCAAGAACC
P97                TCCTACCATT GCCGGATTAG CCGTTTTAGT CTCTAAAACT GAATATGATT ATCAAGAACC
P126-8             TCCTACCATT GCCGGATTAG CCGTTTTAGT CTCTAAAACT GAATATGATT ATCAAGAACC
P127               TCCTACCATT GCCGGATTAG CCGTTTTAGT CTCTAAAACT GAATATGATT ATCAAGAACC
P128-R             TCCTACCATT GCCGGATTAG CCGTTTTAGT CTCTAAAACT GAATATGATT ATCAAGAACC
O213               TCCTACCATT GCCGGATTAG CCGTTTTAGT CTCTAAAACT GAATATGATT ATCAAGAACC
O226               TCCTACCATT GCCGGATTAG CCGTTTTAGT CTCTAAAACT GAATATGATT ATCAAGAACC
Cons-Plank98,7%    TCCTACCATT GCCGGATTAG CCGTTTTAGT CTCTAAAACT GAATATGATT ATCAAGAACC
Nod-HEMclone       TCCCACAATC GCCGGATTAG CAGCTTTAGT ATCTAAAACC CAATATGACT ATCAAGAACC
Nod-BY1clone       TCCCACAATC GCCGGATTAG CAGCTTTAGT ATCTAAAACC CAATATGACT ATCAAGAACC
Nod-F81clone       TCCCACAATC GCCGGATTAG CAGCTTTAGT ATCTAAAACC CAATATGACT ATCAAGAACC
Cons-Nod99,4%      TCCCACAATC GCCGGATTAG CAGCTTTAGT ATCTAAAACC CAATATGACT ATCAAGAACC
Nostoc152          TCCCACCATA GCCGGATTAG CAGCTTTAGT AGCTAAAACC CAATACGATT ATCAAGAACC

                   ....|....| ....|....| ....|....| ....|....| ....|....| ..  SEQ ID NO
                        725        735        745        755        765
Ana90-Leo          CATACCAGCA ATAACTCAAC AAGAATCTTA TCCCATGTCT CATGGACAAC GC    1
Ana299B            CATACCAGCA ATAACTCAAC AAGAATCTTA TCCCATGTCT CATGGACAAC GC    2
Ana315             CATACCAGCA ATAACTCAAC AAGAATCTTA TCCCATGTCT CATGGGCAAC GC    3
AnaPH256           CATACCAGCA ATAACTCAAC AAGAATCTTA TCCCATGTCT CATGGACAAC GC    4
Ana318             CATACCAGCG ATAACTCCAC AAGCATCTTA TCCCATGTCT CATGGACAAC GC    5
Ana202A1           CATACCAGCA ATAACTCAAC AAGAATCTTA TCCCATGTCT CATGGACAAC GC    6
Ana202A2-41        CATACCAGCA ATAACTCAAC AAGAATCTTA TCCCATGTCT CATGGACAAC GC    7
Ana66A             CATACCAGCG ATAGCTCCAC AAGCATCTTA TCCCATGTCT CATGGGCAAC GC    8
Ana83-1            CATACCAGCA ATAACTCAAC AAGAATCTTA TCCCATGTCT CATGGACAAC GC    9
Cons-Ana98%        CATACCAGCA ATAACTCAAC AAGAATCTTA TCCCATGTCT CATGGACAAC GC    10
Mpcc7806-Tillet    CATTTCGGCA ATTCCCCCCC AAAAATCCTA TCCGATGTCT CATGGTCAGC GT    11
M205               CATTTCGGCA ATTCCCCCCC AAAAATCTTA TCCGATGTCT CATGGTCAGC GT    12
M98                CATTTCGGCA ATTCCCCCCC AAAAATCTTA TCCGATGTCT CATGGTCAGC GT    13
Mgl260735          CATTTCGGCA ATTCCCCCCC AAAAATCTTA TCCGATGTCT CATGGTCAGC GT    14
Mgl280646          CATTTCGGCA ATTCCCCCCC AAAAATCTTA TCCGATGTCT CATGGTCAGC GT    15
Mpcc7941           CATTTCGGCA ATTCCCCCCC AAAAATCTTA TCCGATGTCT CATGGTCAGC GT    16
Miz25              CATTTCGGCA ATTCCCCCCC AAAAATCCTA TCCGATGTCT CATGGTCAGC GT    17
MizM5              CATTTCGGCA ATTCCCCCCC AAAAATCTTA TCCGATGTCT CATGGTCAGC GT    18
Mnies102           CATTTCGGCA ATTCCCCCCC AAAAATCTTA TCCGATGTCT CATGGTCAGC GT    19
MniesA89           CATTTCGGCA ATTCCCCCCC AAAAATCCTA TCCGATGTCT CATGGTCAGC GT    20
Cons-Mic97,4%      CATTTCGGCA ATTCCCCCCC AAAAATCTTA TCCGATGTCT CATGGTCAGC GT    21
P49                CATTCCCACA ATTCCTCTGC AAAAATCCTA TCCGATGTCC CATGGGCAAC GT    22
P97                CATTCCCACA ATTCCTCTGC AAAAATCCTA TCCGATGTCC CATGGGCAAC GT    23
P126-8             CATCCCCACA ATTCCTCTGC AAAAATCCTA TCCGATGTCC CATGGGCAAC GT    24
P127               CATCCCCACA ATTCCTCTGC AAAAATCCTA TCCGATGTCC CATGGGCAAC GT    25
P128-R             CATCCCCACA ATTCCTCTGC AAAAATCCTA TCCGATGTCC CATGGGCAAC GT    26
O213               CATTCCCACA ATTCCTCTGC AAAAATCCTA TCCGATGTCC CATGGGCAAC GT    27
O226               CATTCCCACA ATTCCTCTGC AAAAATCCTA TCCGATGTCC CATGGGCAAC GT    28
Cons-Plank98,7%    CATTCCCACA ATTCCTCTGC AAAAATCCTA TCCGATGTCC CATGGGCAAC GT    29
Nod-HEMclone       CATACCAGCA ATAACTCAGC AAACGTCTTA TCCTATGTCT CATGGGCAAC GC    30
Nod-BY1clone       CATACCAGCA ATAACTCAGC AAACGTCTTA TCCTATGTCT CATGGGCAAC GC    31
Nod-F81clone       CATACCAGCA ATAACTCAGC AAACGTCTTA TCCTATGTCT CATGGGCAAC GC    32
Cons-Nod99,4%      CATACCAGCA ATAACTCAGC AAACGTCTTA TCCTATGTCT CATGGGCAAC GC    33
Nostoc152          CATACCTGCA ATAATTCAGC AAAAATCTTA TCCCATGTCT CATGGGCAAC GC    34
```

List of the group-specific probes and their correspondent Zip codes and complementary zip codes.

Polymorphism position	Group name	Zip code	Discriminating probe	SEQ ID NO:	Common probe	SEQ ID NO:
203	ANABAENA MCY3	ZIP 39	ACCAAATTAGCCGCTATGCCG	40	AGATTGAGAAAGCAATTGTCTTACCTATAGAGG	46
594	MICROCYSTIS MCY3	ZIP 40	TCTACTCTTATCAAGGGTTGTAACCCGG	41	GTTCATAAAGAACTAAATGTATCCGTAAAATTRGCTG	47
424	NODULARIA MCY3	ZIP 41	ATTTGCGATCCCTGGCTGAAT	42	TCAAGGGAATAGGTAACTTAACACAGTTAGCG	48
576	NOSTOC MCY3	ZIP 42	ACTTCTTTGAAATTGGTGGACACTCG	43	CTGCTGCTTTCGAGAGTGGTAACTTACG	49
90	Oscillatoria/Planktothrix MCY3	ZIP 2	GATGTTTAGAACCGGGGATTTAGGG	44	AAACAAACTGCTCCCGGTGTGA	50
217	Oscillatoria/Planktothrix MCY3	ZIP 44	AACTTAATCGTTATCCTCAGATTGAGAGAGCT	45	ATTATTCTACCGATATCAGTCAATAATCAAACTCAA	51

Polymorphism position	Group name	Zip code	Zip code sequence	SEQ ID NO:	C-zip code sequence	SEQ ID NO:
203	ANABAENA MCY3	ZIP 39	CAGCCGCGGTACTGAATGCGATGCT	52	AGCATCGCATTCAGTACCGCGGCTG	58
594	MICROCYSTIS MCY3	ZIP 40	CCCCGGATAGCTGACGAGGCTTACG	53	CGTAAGCCTCGTCAGCTATCCGGGG	59
424	NODULARIA MCY3	ZIP 41	TCCGGACAGGTTGGGGTGCGTTTGG	54	CCAAACGCACCCCAACCTGTCCGGA	60
576	NOSTOC MCY3	ZIP 42	CGTAGAGCAACGCGATACCCCCGAC	55	GTCGGGGGTATCGCGTTGCTCTACGG	61
90	Oscillatoria/Planktothrix MCY3	ZIP 2	CGGTCGACGAGCTGCCGCGCAAGAT	56	ATCTTGCGCGGCAGCTCGTCGACCG	62
217	Oscillatoria/Planktothrix MCY3	ZIP 44	AGCAGCAGTGACAATGCCACCGCCG	57	CGGCGGTGGCATTGTCACTGCTGCT	63

Fig. 20

SEQ ID NO 67

Anabaena *mcyE* amino acid sequence

MAQNTDYKKLIATTLTKMEAMQARITELETRQSEPIAVVGMGCRFPGGISSPEAYWNFCQ
AGLDAIVEVPQSRWDISKFYAPEPTPGKMNTRYGGFLQQDITEFDARFFSISSREATSMD
PQHRLLLEVAWEALENANLPPTNLAGDRVGVFVGITSVDHAMTVYKSKYDEIDSFFGTGN
SLSAAAGRLSYFLNLRGPCMSIDAACASSLVALHQAIRSLRNHECKIALVGGVNLILDPA
ITINLCQSGMMSPDGRCKTFDAAANGYVRGEGCGVLVLKRLSVAEKNGNRILALLRGSAV
NHNGAAAGLTVPSGPAQQDLLRQALADARVKPQEVGYIEAHGTGTSLGDPIEMNAIAAVY
GERSQPLYVGSVKTNIGHLEAAAGIAGTIKTILALQHGEIPSHLHFQEPNPLINWQGYPI
KIPSQAIPWSNNGQVRIAGVSSFGFSGTNAHIIIEQAPAANIPEIKLQRPSHLLTISAHS
ETGLKELALRFHTRLESHPEMGDICHSAAIGRSSLPERLAIVADTLTELQQRLAAFAEEK
NIDHGIFYRRFTGEKYPKIVFLFTGQGACYAGMGNQLYQTQPTFRQYIDQCADILGNYLE
FPLQQILFGDRTDLLNQTAYAQPAIFALEYSLAMLWQSWGIKPSLLIGHSVGEYVAACIA
GVFSLEAGLALIVKRGQLMQTAPLGKMASVFADEATVSALIQNYGNTVSIAAINHPQQIV
ISGESNSIDEIVANCKSQKIAVQLLSVNGAFHSPLMESILDDFEIAARFVSYHPPQILLV
SGIDGQPLTTAPDASYWRQQSRQAVQYFQSLITALNKGYNLFLEVGPRPILAEQGRRYND
DAIWLSSLNRGLDNWQTMLSALAQLYINGVNFNAEKFNKDYGYRNIQLPNYPFQRKRFQF
KSTVLSQSSLTKEVPLERELMETNMNLAKVANIKKNQQEIGNKLKSILALLLKEDENDIR
DDETLLNLGADSIILTDFVRKIEEKFGVKVKIDQLFTDLQTIDEISIYLSDYIKQKPSNT
SDETAINDILTKTSVQVSNSESELNNYLWVISQLQPIAVAYILKALEVLGKRLSIADTWT
TEDLLQTLPIASKYHILVNRYLKTLEQTGIIQNQGNVWIVKSLPTPFSLPEAIENLQTIC
PAAKPELDMLQRCGENLAEVLKGNIDPLELIFPAGSVVHAESIYGNSPVSRLMNQRVSQA
INSILNNFSSSDRPYQIIEVGGGTGATSEAIVNNLNLNHTTYFFTELSPVLLNKARQKFK
NRHKFNFNQLDIEKSPVSQGLTAHSYHIVVAANVLHSTRNITETLNNIRELLIPGGYLVL
LETVENNSWLDLTFGLTPGWWRFQDKELRLDTPLLSGESWCAALKRCGFVNADIYSQQNN
ISIYNGQELIIASTSPESAIDHQSKTVAVSIPTSGKEALMMAQLQSLKELKDIHEKTIIK
QLEILQSAPVAPSNTPEVLLIQTETAPTPKISKTETTPPTQKISSPNLNPLALKLTESKS
LTEQQQAFIQKLEIVYNQRTAKSKAYSQNSRKTMVDVKPTIDFRMALKEFQYPIVSESAQ
GAYFRDIDGNDYIDLAMGFGVNFFGHSPDFVLTEIQQQMQHGIGLGMQSNIAAETAALIC
EITGVERVAFSNTGTEAIMAAVRIARSRTKRQKIVIFAGSYHGTFDGILARAGEEAGTAE
PLSLGTPSGMVEDVIVLTYGAEESLEIIAEQADNLAAVLVEPVQSRKPDLQPKEFIQKLR
KLTQQKEIALIFDEIITGFRITPGGAQEWFEIEADIVVYGKAIGGGLPISMICGKADFLD
TVDGGFWSYGDDSHPQTELTAYGGTFCRHPLALAACRAVLLHLREQGATLQETVNQLTNR
LAIEVNQFFQETGIPIRIVHFGSLFRFESSGAYSIFLKPIELPLFYYLLNLKGVYTWEKR
VCFLSTRHTNEDINKVVAAVKEAIIELRQAGFFANAKPPQTKKREASDRTEDEDARNNLN
QQFPTSEAQRQLWLLAELDTTASASYNVTTSLELRGALDISSLQQAINEVVNRHEALRTK
ILEQGELQEVISSVTIDLPLINLMDEDNPEATALVLRTELSQKPFDLGVAPLFAAVLMRL
APEHYLLTLKTHHIVADGWSLGLILNELGKLYSAKIGVATESLSPPMQFRKYLALRQQEA
QSPQMQAHRDFWLKTYEGEIPIFELPTDFPRPAVKTYTGGRESKIIAPQLWQNLQTVGRK
NQATLFMTMFAAYTAFLRRISGHDDLVIGIPISGRQVEGSEKLVGFCSQFLPIRIQTDVT
ASFVTHLRHTKETLIAAFKHQTHALEELLAALQLQRDFSRSPLISVSFNLDPKLTLPEFE
GLNVSLPPEPIGYTPFDLGFNFIEVNDALIIYCNYNTELFKPETIKQFLESFEILMQGVI
KDANILLSELHLLTQVQQEELLAKLTGSTIELPQNSTIIDDFIAQVKSTPDAPALIVEEK
TLTYRELNEKVNRLTNYLREKYNLGAGKAIALAIGRNQNLIIAILATFKTGAIYVPIDPQ
YPSSRIDFILKDSGCHLCLTESNFISQLPQEIEAICLDKIDNILTDFDINEPNFQPDTNQ
IAYILYTSGSTGNPKGVMGRHISILNVIRSLRLTFNLNKHPEWRYIFTAPVTHDPSFRNI
FLPLTIGAALYMYEVQHIGHLVSFLQENKINVLHTTPSIYRFILAVLAPEETIPTLKYIS
CGGEKLDRETAIALRKRFPAEIVSNVYGSTETCVGVSQYTIDDNLNTDVPLGQVFHNNRL
FVLDEFNHPVPLHVIGEICVEGAALAVGYRNLPEITREKFQPNFLNSEKILFRTGDLGKQ
IAPGVIEFIGRKDNQVKVNGYRVDPGEIEYQISRYAEIEKAIVLPIEVNNQIQLSAYCQT
DKDIKFSEIREFLAKYLPVYMIPSSFIFLKQFPLTKHGKFDLRSLIALKPTDQLTQVSYT
APRNTLESKLVHIWEKILTKHPIGIFDNFFEIGGHSLLLSRVVTHVHKELNVLVKLADFF
KVPTILGLAALISKAQSNYQEPIPAITQQESYPMSHGQRRLWALEFLDHNHYAYGMPSAY
QFNGDLNIAAFENAFKKLIERHEILRTTFTLINNEPRQIVNEQVDFAVNQIDLVDDENQA
AKIAEAIRNNAKTTFDLESGLLLKINVLKLSQQSNIVLFNMHHIISDGWSAGVLIKDFLA
HYHAYGKENVELPPPLRIHYKDYTSWQNQQLQTPKLQAQRDYWLPKLIPAPAPLDLPLDY
TRPAVQSFSGSVVIWKPNQEFIKDFELLTKTQEASLFMGLLTLVKGFLFRYTEQNEITVG
SPIAGRNHPDLEEQIGFYVNTLVLRDQITVDDSFATLLAKVKTTTIEAYDNQEYPFDKLV
SDLNFKRDPSRNPLFDVVVVLQNNQNVDLAIDGIAVNTLEQELVTAKFDLEFIFVDEAEL
YLKLIYNTDIFANERISLMIKLLETLLEEVVKSPDTPLLHLCDHTDKACQEDNSLFATNF
NF

Fig. 21

Anabaena mcyE Length: 10449 SEQ ID NO 68

```
1 ATGGCTCAAA ATACAGATTA TAAAAAATTA ATAGCTACAA CCCTCACAAA
51 AATGGAGGCG ATGCAGGCTC GCATTACTGA ATTAGAAACT AGACAAAGTG
101 AGCCGATTGC TGTTGTGGGA ATGGGTTGTC GTTTCCCTGG AGGAATCAGT
151 TCTCCTGAAG CTTATTGGAA CTTTTGTCAA GCTGGACTTG ATGCAATTGT
201 AGAAGTTCCT CAAAGCCGTT GGGATATCTC AAAATTTTAT GCTCCAGAGC
251 CTACTCCTGG CAAAATGAAC ACTCGTTATG GGGGATTTTT ACAACAGGAT
301 ATTACAGAAT TTGATGCCCG TTTCTTCTCC ATATCTTCCC GCGAAGCAAC
351 TTCAATGGAT CCTCAACACA GGTTATTACT TGAGGTGGCG TGGGAAGCGT
401 TAGAAAACGC CAATTTACCA CCAACTAATT TAGCAGGCGA TCGCGTGGGT
451 GTGTTTGTCG GTATCACTAG TGTTGACCAC GCGATGACAG TCTACAAAAG
501 CAAGTATGAT GAAATCGATT CTTTTTTTGG TACAGGAAAC TCCCTAAGTG
551 CAGCAGCAGG TAGGTTATCT TATTTTCTCA ACCTCCGCGG ACCTTGTATG
601 TCTATTGATG CAGCCTGTGC TTCTTCATTG GTTGCACTTC ACCAAGCTAT
651 TCGCAGCTTG AGAAATCATG AGTGCAAAAT AGCCTTAGTA GGTGGTGTCA
701 ATCTCATCTT AGATCCGGCA ATTACGATTA ACCTTTGTCA GTCGGGGATG
751 ATGTCTCCCG ATGGTCGCTG CAAGACTTTT GATGCGGCTG CAAATGGATA
801 TGTGCGGGGC GAAGGATGCG GGGTTTTAGT TCTCAAACGC CTGTCTGTCG
851 CTGAAAAAAA TGGCAATCGC ATTCTCGCAT TACTACGGGG GTCTGCGGTC
901 AATCATAACG GTGCAGCCGC CGGTTTAACA GTTCCCAGTG GCCCCGCCCA
951 ACAAGATTTA CTTCGTCAAG CCTTAGCTGA TGCTAGAGTC AAACCCCAGG
1001 AAGTCGGTTA TATAGAAGCA CATGGTACGG GTACTTCTTT AGGCGACCCC
1051 ATTGAGATGA ACGCGATCGC TGCTGTGTAT GGAGAGCGAT CGCAACCTCT
1101 TTACGTCGGT TCAGTTAAGA CTAATATCGG ACATCTCGAA GCTGCCGCAG
1151 GGATAGCAGG CACAATCAAG ACAATTCTCG CCCTCCAGCA TGGTGAGATT
1201 CCGTCTCATC TCCACTTCCA AGAACCCAAT CCCCTGATTA ATTGGCAAGG
1251 ATACCCGATC AAAATTCCTA GTCAAGCCAT ACCCTGGTCG AACAATGGCC
1301 AAGTCCGTAT CGCTGGAGTC AGTTCCTTCG GTTTTTCTGG AACCAATGCT
1351 CACATAATTA TTGAACAAGC ACCTGCTGCC AACATACCAG AAATTAAACT
1401 GCAACGTCCC AGCCATCTGT TGACGATTTC TGCTCATAGC GAAACGGGTT
1451 TAAAAGAACT AGCACTACGT TTTCACACCC GCTTAGAGTC TCATCCAGAG
1501 ATGGGAGATA TTTGTCATAG TGCGGCAATT GGTAGGTCGT CTTTACCTGA
1551 ACGTTTAGCG ATCGTTGCAG ATACATTGAC AGAGTTGCAA CAAAGATTAG
1601 CAGCTTTTGC TGAGGAAAAA AATATTGATC ATGGGATTTT TTATCGACGT
1651 TTTACAGGAG AAAAATACCC TAAAATAGTC TTTCTTTTCA CTGGTCAAGG
1701 AGCTTGTTAT GCGGGTATGG GTAATCAACT TTACCAAACC CAACCGACAT
1751 TTCGCCAATA TATTGATCAA TGTGCAGACA TTTTAGGAAA TTATTTAGAG
1801 TTTCCACTAC AACAAATATT ATTTGGCGAT CGCACAGATT TACTCAATCA
1851 AACTGCATAC GCTCAACCAG CTATATTTGC TCTAGAATAT TCACTTGCCA
1901 TGTTGTGGCA ATCTTGGGGA ATCAAACCTA GTTTACTAAT TGGCCATAGC
1951 GTTGGTGAAT ATGTTGCAGC TTGTATTGCA GGAGTATTTA GTTTAGAAGC
2001 AGGTCTAGCA TTAATTGTTA AACGTGGGCA ATTAATGCAA ACTGCACCTC
2051 TAGGAAAAAT GGCTTCAGTT TTTGCAGATG AAGCTACGGT ATCTGCTCTC
2101 ATTCAAAACT ATGGTAATAC AGTTTCCATC GCGGCTATTA ATCATCCTCA
2151 ACAAATTGTG ATTTCTGGGG AAAGTAATAG TATTGATGAG ATTGTCGCCA
2201 ACTGCAAAAG CCAAAAAATA GCCGTTCAAT TACTATCTGT AAATGGGGCT
2251 TTTCATTCTC CATTAATGGA GTCAATCTTA GATGATTTTG AAATTGCTGC
2301 TCGTGAGGTT TCGTATCATC CTCCTCAAAT TCTACTAGTT TCAGGAATTG
2351 ATGGACAACC TTTAACAACT GCGCCTGATG CAAGTTACTG GAGACAACAA
2401 AGCCGTCAAG CAGTGCAGTA TTTCCAGAGT CTAATTACGG CGTTAAACAA
2451 AGGATATAAC CTATTTTTAG AAGTGGGTCC AAGACCAATA CTAGCAGAAC
2501 AAGGCCGTCG TTATAACGAT GACGCTATAT GGCTTAGTTC TCTCAATAGA
2551 GGCTTGGATA ACTGGCAGAC GATGCTATCA GCACTTGCCC AACTCTACAT
2601 TAATGGAGTT AATTTTAACG CCGAAAAATT CAATAAAGAT TATGGCTATA
2651 GAAATATCCA ATTACCAAAT TATCCTTTTC AAAGAAAACG TTTTCAATTT
2701 AAATCTACTG TTTTATCACA ATCAAGTTTG ACCAAAGAAG TGCCTTTAGA
2751 AAGAGAGCTA ATGGAGACTA ATATGAATTT AGCAAAAGTT GCAAATATCA
2801 AAAAAAATCA GCAAGAAATC GGTAATAAAT TAAAGTCTAT CTTGGCTTTG
2851 CTACTTAAAG AAGATGAAAA TGACATTAGA GATGATGAGA CATTATTAAA
2901 TCTGGGAGCA GATTCGATTA TTTTGACCGA CTTTGTTCGC AAAATTGAAG
```

Fig. 22A

```
2951 AAAAATTTGG CGTAAAAGTC AAGATAGATC AATTATTTAC TGATTTACAA
3001 ACAATTGACG AAATATCTAT TTATCTATCT GATTATATTA AACAAAAACC
3051 ATCAAATACA TCTGATGAAA CAGCAATAAA TGATATTTTA ACTAAAACCT
3101 CAGTTCAAGT ATCTAATTCC GAGTCAGAAT TAAATAATTA TCTGTGGGTG
3151 ATTTCCCAAT TACAACCAAT TGCTGTTGCT TATATCTTGA AAGCTTTAGA
3201 AGTATTAGGT AAAAGACTAA GTATAGCAGA CACCTGGACA ACAGAGGATT
3251 TATTACAAAC CTTACCCATA GCCTCAAAAT ATCATATTTT AGTTAATCGT
3301 TATTTAAAAA CTTTAGAGCA AACTGGAATT ATTCAAAATC AAGGTAATGT
3351 ATGGATAGTA AAAAGCCTAC CTACGCCTTT TTCTTTACCA GAAGCTATAG
3401 AGAATCTGCA AACAATTTGT CCAGCAGCAA AACCTGAATT GGATATGCTA
3451 CAACGTTGTG GAGAAAATTT AGCAGAAGTT CTCAAAGGAA ATATAGATCC
3501 TTTAGAGTTA ATTTTCCCAG CAGGTTCAGT TGTCCATGCA GAAAGTATAT
3551 ATGGCAATTC TCCTGTATCT CGTTTGATGA ATCAACGAGT ATCCCAAGCA
3601 ATTAACTCTA TTTTAAATAA TTTTTCTAGT AGCGATCGCC CTTACCAAAT
3651 TATTGAAGTG GGAGGTGGTA CAGGTGCGAC ATCTGAAGCA ATTGTTAATA
3701 ATTTAAATCT CAATCATACA ACTTATTTTT TCACTGAACT TTCTCCTGTT
3751 CTCCTCAATA AAGCACGGCA AAAATTTAAA AATAGGCATA AGTTTAACTT
3801 CAATCAATTA GATATTGAAA AATCACCAGT TTCTCAAGGG TTAACAGCGC
3851 ATTCCTATCA CATAGTTGTT GCTGCTAACG TGCTTCATAG TACCCGAAAT
3901 ATTACAGAAA CACTCAATAA TATTCGAGAA CTACTAATAC CTGGTGGTTA
3951 CTTGGTATTA CTAGAAACAG TAGAAAATAA TTCATGGCTT GATTTAACTT
4001 TTGGACTCAC ACCTGGATGG TGGCGTTTTC AAGATAAAGA GTTACGTTTA
4051 GATACTCCAT TGCTAAGTGG AGAAAGTTGG TGTGCTGCTT TAAAACGTTG
4101 TGGATTTGTC AATGCAGATA TTTATTCTCA ACAAAATAAT ATCAGCATCT
4151 ACAATGGACA AGAGTTAATT ATTGCTTCCA CTTCTCCTGA GAGTGCTATT
4201 GATCACCAAT CAAAAACAGT AGCTGTATCA ATACCAACCT CTGGCAAAGA
4251 AGCTTTGATG ATGGCTCAAT TGCAATCTTT GAAAGAACTC AAAGATATTC
4301 ATGAAAAAAC TATTATCAAG CAATTAGAAA TACTGCAATC TGCACCTGTT
4351 GCACCAAGCA ATACACCAGA AGTTTTGTTG ATTCAAACAG AAACTGCGCC
4401 TACTCCTAAA ATAAGCAAAA CAGAAACCAC ACCTCCTACC CAAAAAATTA
4451 GTTCTCCAAA TTTGAATCCA CTAGCTTTAA AACTGACAGA GAGTAAATCT
4501 TTAACCGAAC AACAGCAAGC TTTCATCCAA AAATTAGAGA TTGTTTATAA
4551 TCAAAGAACA GCAAAATCCA AGGCATATTC GCAAAACAGC CGGAAGACAA
4601 TGGTTGATGT GAAACCTACT ATTGACTTCC GTATGGCTTT AAAAGAATTT
4651 CAGTATCCCA TCGTTTCAGA ATCAGCCCAA GGTGCATATT TTCGTGATAT
4701 CGATGGCAAT GATTACATCG ACTTAGCAAT GGGGTTTGGA GTTAACTTTT
4751 TTGGGCATAG TCCTGATTTT GTGTTGACAG AAATTCAGCA ACAAATGCAA
4801 CACGGAATTG GTTTAGGAAT GCAGTCTAAT ATTGCAGCAG AAACAGCTGC
4851 TTTAATTTGT GAAATAACAG GTGTAGAACG AGTCGCTTTT AGTAATACTG
4901 GTACAGAAGC GATTATGGCA GCAGTTCGTA TTGCTCGTTC TCGGACAAAA
4951 CGCCAAAAAA TTGTGATTTT TGCTGGTTCT TATCACGGTA CTTTTGATGG
5001 CATTTTAGCC CGCGCTGGCG AGGAAGCAGG AACAGCTGAA CCATTGAGTC
5051 TTGGTACACC ATCAGGAATG GTTGAAGATG TTATAGTTCT TACCTATGGA
5101 GCGGAAGAAA GTCTAGAAAT AATTGCAGAG CAAGCTGATA ATTTAGCAGC
5151 AGTTTTAGTA GAACCTGTGC AGAGCCGTAA ACCAGATTTA CAGCCTAAAG
5201 AATTTATCCA AAAATTACGT AAATTGACTC AACAAAAAGA AATTGCCCTA
5251 ATTTTTGATG AGATTATTAC TGGATTTCGC ATTACTCCAG GAGGAGCGCA
5301 GGAATGGTTT GAGATTGAGG CAGATATAGT AGTTTATGGT AAGGCGATCG
5351 GTGGTGGTTT GCCTATCAGT ATGATCTGCG GTAAAGCAGA TTTCTTAGAC
5401 ACAGTTGATG GCGGTTTCTG GAGTTATGGA GATGATTCCC ATCCGCAAAC
5451 AGAATTAACT GCTTATGGTG GTACTTTCTG TCGTCATCCC TTAGCTTTGG
5501 CTGCTTGTCG GGCAGTGCTT TTACATTTAC GTGAACAAGG CGCAACGCTT
5551 CAAGAAACAG TCAATCAATT AACTAATCGA TTAGCTATTG AGGTTAATCA
5601 GTTCTTTCAA GAAACAGGAA TTCCTATCCG CATTGTTCAT TTTGGCTCTT
5651 TGTTTCGCTT TGAATCTTCT GGTGCTTATA GTATTTTCCT CAAACCTATC
5701 GAATTACCAC TTTTTTACTA CTTATTAAAT CTCAAAGGTG TGTATACATG
5751 GGAAAAACGA GTTTGTTTTC TCTCAACTCG CCATACCAAT GAGGATATAA
5801 ATAAAGTAGT AGCTGCTGTC AAAGAAGCGA TTATAGAACT GCGACAAGCA
5851 GGTTTTTTCG CAAATGCTAA ACCACCACAA ACAAAGAAAA GAGAAGCAAG
5901 CGATCGCACT GAAGATGAAG ATGCTAGGAA TAATTTAAAT CAGCAATTTC
5951 CTACCAGTGA AGCGCAACGT CAGCTTTGGC TATTGGCAGA ATTAGATACA
6001 ACCGCTTCAG CATCATATAA TGTGACTACT TCCTTAGAAT TGCGTGGTGC
6051 TTTAGATATT TCATCTTTAC AACAAGCGAT TAATGAAGTA GTCAATCGCC
```

Fig. 22B

```
6101 ACGAAGCGTT GCGAACCAAA ATATTAGAAC AAGGCGAACT ACAGGAAGTA
6151 ATCAGTAGTG TCACAATTGA CTTACCGTTA ATCAATTTGA TGGATGAAGA
6201 TAACCCAGAA GCAACTGCTT TAGTATTGAG AACTGAGTTA TCGCAAAAGC
6251 CTTTTGACTT GGGTGTTGCT CCTTTATTTG CGGCTGTGCT GATGCGTTTA
6301 GCACCTGAAC ATTATCTGCT AACTTTAAAA ACTCATCATA TTGTTGCAGA
6351 TGGTTGGTCA CTAGGGCTAA TTTTGAATGA ACTCGGCAAA CTTTATTCAG
6401 CAAAAATTGG TGTTGCTACA GAATCTTTAT CGCCACCAAT GCAGTTTCGC
6451 AAATATTTAG CTTTGCGACA GCAAGAAGCG CAAAGTCCAC AAATGCAAGC
6501 ACATCGTGAT TTTTGGTTAA AAACTTATGA GGGAGAAATC CCCATATTTG
6551 AACTTCCTAC AGACTTTCCT CGTCCTGCTG TCAAAACTTA CACAGGCGGT
6601 AGAGAAAGTA AAATTATTGC TCCTCAACTA TGGCAAAATT TACAGACTGT
6651 AGGACGCAAA AATCAAGCAA CATTATTTAT GACAATGTTT GCTGCTTACA
6701 CAGCTTTTTT GCGGCGCATT TCCGGTCATG ATGATTTAGT AATTGGTATT
6751 CCCATTTCTG GACGACAAGT CGAAGGAAGC GAAAAATTAG TTGGCTTTTG
6801 TTCGCAATTT TTACCGATTC GTATCCAGAC AGATGTCACT GCTTCTTTTG
6851 TTACACATCT TCGTCACACC AAAGAAACAC TCATAGCCGC TTTTAAACAC
6901 CAAACTCACG CCTTAGAAGA ATTATTAGCA GCCTTACAAT TACAACGAGA
6951 TTTTTCGCGT TCTCCCCTGA TTTCAGTCTC TTTCAATCTC GACCCTAAAT
7001 TAACTTTACC TGAATTTGAA GGACTGAACG TATCGCTACC ACCAGAGCCA
7051 ATTGGTTACA CTCCTTTCGA TTTGGGCTTT AATTTCATCG AAGTCAATGA
7101 CGCACTGATT ATCTACTGCA ACTACAATAC AGAACTGTTT AAGCCGGAAA
7151 CTATCAAACA GTTTCTAGAA AGTTTTGAAA TTTTGATGCA GGGTGTAATT
7201 AAAGATGCCA ATATTTTACT GTCTGAGTTG CATTTATTAA CACAAGTGCA
7251 ACAGGAAGAA TTATTGGCAA AATTAACAGG TTCAACAATA GAATTACCCC
7301 AAAATTCTAC AATCATTGAT GACTTTATTG CTCAAGTTAA ATCTACACCC
7351 GATGCACCTG CATTAATAGT TGAGGAAAAG ACTCTTACCT ATCGAGAATT
7401 GAATGAGAAG GTAAATCGTT TAACCAATTA CTTACGTGAG AAATATAACC
7451 TTGGTGCGGG GAAGGCGATC GCACTCGCAA TCGGACGCAA TCAAAACTTA
7501 ATTATCGCCA TTCTGGCTAC TTTTAAAACC GGGGCGATAT ATGTGCCGAT
7551 AGACCCCCAA TATCCTAGCA GTCGCATCGA TTTTATTTTA AAAGATAGTG
7601 GCTGTCATCT CTGTCTTACC GAGAGTAATT TTATCTCCCA ATTACCCCAA
7651 GAAATCGAAG CAATCTGCTT AGATAAAATC GACAATATTT TGACGGATTT
7701 TGACATAAAT GAGCCTAATT TCCAACCAGA CACTAACCAA ATTGCTTATA
7751 TTTTATACAC TTCTGGCTCT ACCGGAAACC CTAAAGGAGT CATGGGTCGT
7801 CACATTTCTA TTTTGAATGT AATTCGGAGT TTACGACTGA CATTTAACTT
7851 AAATAAACAT CCTGAATGGC GTTATATTTT CACTGCACCT GTCACCCATG
7901 ACCCATCTTT TCGTAATATT TTCCTACCTT TAACCATAGG TGCTGCTTTA
7951 TATATGTACG AAGTACAGCA TATAGGGCAT TTAGTTTCAT TTTTACAAGA
8001 AAACAAAATC AATGTACTGC ACACAACACC ATCTATTTAC CGAGAAATTT
8051 TAGCTGTACT CGCACCCGAA GAAACTATCC CCACTTTAAA ATATATCTCT
8101 TGCGGGGGAG AAAAATTAGA CCGAGAAACT GCGATCGCTT TACGAAAACG
8151 CTTCCCTGCT GAAATTGTCA GCAATGTCTA CGGTTCTACA GAAACTTGTG
8201 TAGGAGTATC TCAATATACA ATAGACGACA ACTTAAATAC AGACGTACCT
8251 CTTGGTCAGG TTTTCCATAA CAATCGATTA TTTGTCTTAG ATGAATTCAA
8301 TCACCCCGTA CCTCTGCACG TAATCGGTGA AATTTGCGTA GAAGGTGCAG
8351 CATTAGCAGT TGGTTATCGT AATCTTCCTG AAATTACCAG AGAAAAATTT
8401 CAACCCAACT TTTTAAATTC AGAAAAAATT CTCTTTAGAA CGGGAGATTT
8451 AGGTAAACAA ATTGCTCCAG GTGTGATTGA ATTTATAGGG CGAAAAGATA
8501 ATCAAGTTAA GGTCAATGGC TATCGTGTAG ATCCAGGAGA AATTGAATAC
8551 CAAATTAGCC GCTATGCCGA GATTGAGAAA GCAATTGTCT TACCTATAGA
8601 GGTAAATAAC CAAATTCAAT TATCTGCTTA TTGTCAAACT GATAAAGATA
8651 TAAAATTTTC TGAAATCAGA GAATTTTTAG CTAAATATTT GCCAGTTTAC
8701 ATGATTCCTA GTTCCTTTAT CTTCTTAAAG CAATTTCCCT TAACTAAACA
8751 TGGCAAATTT GACTTGCGAT CGCTCATCGC TCTCAAGCCA ACAGATCAAT
8801 TAACACAAGT CTCTTATACT GCACCGCGTA ATACTTTAGA ATCAAAGCTA
8851 GTCCATATTT GGGAAAAAAT TCTCACTAAA CATCCCATTG GAATTTTTGA
8901 TAACTTTTTT GAAATCGGCG GACACTCTCT GCTCCTTTCT AGAGTAGTCA
8951 CTCACGTCCA TAAAGAATTA AATGTATTAG TTAAATTAGC TGATTTCTTC
9001 AAAGTTCCCA CAATTCTTGG ATTAGCAGCT TTAATATCTA AAGCTCAATC
9051 TAACTATCAA GAACCCATAC CAGCAATAAC TCAACAAGAA TCTTATCCCA
9101 TGTCTCATGG ACAACGCCGT CTCTGGGCTT TAGAATTCCT CGATCATAAC
9151 CATTATGCTT ACGGAATGCC AAGTGCTTAT CAATTCAATG GTGATTTAAA
9201 TATCGCTGCC TTTGAAAATG CTTTTAAAAA ATTGATAGAA CGCCATGAAA
```

Fig. 22C

```
9251 TTTTACGCAC TACGTTTACT TTAATCAATA ACGAACCTCG TCAAATAGTA
9301 AACGAACAGG TAGATTTTGC AGTCAATCAA ATAGACTTAG TAGATGATGA
9351 AAACCAAGCA GCAAAAATTG CTGAAGCTAT TCGTAACAAT GCCAAAACTA
9401 CTTTTGATTT AGAATCCGGT CTTCTACTCA AGATTAATGT ACTAAAGCTG
9451 AGTCAACAGA GTAATATAGT CCTATTCAAT ATGCACCATA TCATTTCAGA
9501 TGGTTGGTCA GCAGGCGTTT TAATTAAAGA TTTTCTTGCA CATTATCATG
9551 CCTATGGAAA AGAGAATGTA GAATTACCCC CACCTCTAAG AATTCATTAT
9601 AAAGACTATA CTTCATGGCA AAATCAGCAA CTCCAAACAC CAAAATTGCA
9651 AGCACAAAGG GATTACTGGT TACCTAAACT GATTCCTGCT CCAGCACCGC
9701 TTGATTTACC TTTAGATTAT ACTCGACCAG CCGTACAAAG TTTTTCTGGT
9751 TCTGTGGTTA TCTGGAAGCC AAATCAAGAA TTCATCAAGG ATTTTGAGTT
9801 ATTAACTAAG ACTCAAGAAG CTAGTTTATT TATGGGCTTA CTAACATTAG
9851 TCAAAGGTTT TCTGTTTCGC TATACCGAGC AGAATGAAAT TACAGTAGGC
9901 TCTCCCATTG CTGGAAGAAA TCATCCCGAT TTAGAAGAAC AAATAGGTTT
9951 TTATGTTAAT ACTTTAGTCT TGCGCGATCA AATAACAGTA GACGATAGCT
10001 TTGCTACTTT ACTGGCAAAA GTGAAAACAA CCACAATAGA GGCATATGAT
10051 AACCAAGAAT ATCCTTTTGA TAAACTAGTA TCAGATTTAA ATTTTAAACG
10101 CGATCCTAGC CGTAACCCTT TATTTGATGT TGTAGTTGTT CTGCAAAATA
10151 ATCAAAATGT AGACCTAGCA ATAGATGGAA TTGCTGTTAA TACCCTTGAG
10201 CAGGAACTTG TTACTGCTAA ATTTGATCTA GAGTTCATTT TTGTGGACGA
10251 AGCGGAATTA TATTTGAAGT TAATCTACAA CACAGATATA TTCGCTAACG
10301 AACGTATTTC ATTGATGATA AAGTTGTTAG AAACTCTTCT AGAAGAAGTA
10351 GTCAAATCTC CTGACACACC ATTATTACAT CTGTGCGATC ACACAGATAA
10401 AGCCTGTCAA GAAGATAACA GTCTTTTCGC CACAAATTTT AATTTTTAG
```

Fig. 22D

Anabaena mcyD amino acid sequence SEQ ID NO 69

MELEISNDKSVQARVFKALNEAKERLKAVEAQQSEPIAVVGLGCRFGKNISTPEAFWSFL
KAGGNTLTDVPSDRWNVSDFYDSDRSKPGKIYVSQGGFLNDVSMFDAHFFGIAPREALHI
DPQQRLLLEVGYEALEDAGMAAPTARIGKTGVFIGITNNDYARLIAPGGDYSTIGAYHIS
GNHINAAAGRISYLLNLNGPSLAIDTACSSSLVAVHLACRSLRSQECRQALVGGVNLILT
PEVPIALCRNQMLAADGRCKTFDESADGFGIGEGCGVVVLKRLSEALADSDRIWALIRGT
SVNNDGASGGFTVPNGPMQTDLIRDALADARLSANAIDYVEAHGTGTSLGDPIEIKAIAD
ALCVNRSPSQPLLIGSVKTNLGHLAAAAGISGLIKTILSIYHGEIPAHLNFQKPNPHINW
DSIPLHVVTQTRPWTQGNGKIKAAGVSSFGASGTNAHAILSEPPQVQSSTKTTFPVLVTL
SAKDEERLRLLAANFVNYIDQKPEVNLTDIAFSVNTGRFHHKQRLTLLVSNTASFKEQLR
AFVNGNDREANIIYGEVRVNTKIAFVCNDARTIFPAMGKELFVNEPVFRDALMQCDRLFQ
NYLNFSIEDLLDNNQDINDPLYVHPTLFALQYALCELWKSWGIRPSAILGNGLGEYIAAQ
QAGIFSIGDAVKLVATRVHLFSEQSGEKEFLAFQKVAEEVSYRNPQLTLISNNGEVANNY
IKTAAYWLQQLHHLDNIEKGIFTLHQMGYQIFLEIGNQPINNIVHSKLPAKEILCLCSLA
KDCSNYQQIQTVIAELYVRGVNINWQEFHRSCSGQKISLPSSPFIRKRYWVDAIKSNRSG
HPLLGDRFPSPLSSIQYRASISQNNPEFLTEHQVFDKPIFPGAAFIEMALAAAQSQAVIL
ENIEFQKALLLQEQEDTLQLIIEQNCFQIYQQSNNNWDVLVTGGIDSLNILDANRQNLEQ
IAANCPEQMEINSFYERYQKSGVNYGSSFQIINKLQRGKNSAFAKIKLTQIQCLEAEKYK
FHPVMLDACFQAIAAILFQQESSVTYVPVRIAKFQFFKSPGASVISAARLTKNSNNYIIS
NIDIYSEQGELLVAITGFELKSVQSQEIIRHQIQPQSYMEEWITLASLLPDGRDYLLNPD
AIKTLVQPQQLEQQLAEKLYQYERLLEEMETLSVSYIWEGLKELNWQPQLGQIYPEEQIA
TQGGVVDFYRPLLSRCLAILAEEGIITQQKDGWLLAKEPVISSSQLPIQQLRREFPDYLA
EINLIERCGSALAAVMRRQIEPLELLFPQGDLNAIASVYSDAAGAKLMNELVAATIKTVV
ANLPTNRQLRILEIGGGTGSSTAAILPHLPPEQIEYTFTDISSSFLTRAKENFSNYPFIK
YQTLDIEKAPISQGFLPSYFDIIIAANVLHATADINETLNNVRSLLAPNAILILLESTGA
RPWVDLTFGLTEGWWLCSQDPHRNGYPLVDTEHWQNLLAKHQFTEINIIEPTNPKTRNLL
QQSVIIAKSSLPSLCTSVSWREIIFADTNGIARSLITPFQQRGITCSLISPQDINPDNPD
DYLSLLQNLITPETREIIYLWSLQEIEGEIYQAVEIHCRRFLFLLQALLQQENPPALILV
TQGSVPAKEITTLTSPAQSSLLGMALSLVLEHPELNFRAIDLDPHAQDLGEKLFREIHNN
TQENRVALRGEQRFCPRLVERKLADGNINFRQDGFYLISGGTGGLGLATARWMIEHGACH
LVLCSRSGAKALNPEILASLQSINEDIQIKDVDVTDAEKLHALLEECRSQYPLRGIFHIA
GTLDDTTLLRLTPERFNYVLAPKVKGTWLLHQLTLNDTLDFFVCYTSAVSLIGSAGQANA
AAANAFEDAFTYYRHAHNLPATVINWGPWSEIGAAVDRNVLERLAAKGYDAIAPDLALNT
LEKILFNQIVRAGVIAIDWQRFPYINQSFYQNFLPQIKPKSQTASNLLEQWQIIPVKQRR
DLLIRQISLRVCTVLGLSTHEVSPQQGFFDMGMDSLTSTELRNLLQTDFNCSLPTTIAFR
FPNVETLADYLLREILVTSEVQTPVQQLIQEVPQIQIEKYTPEKPQQEEDPIVIVGMACR
FPGGANDLESFWQLLEQGKDAVREIPSDRWDMQAWYHPDPDTPGKIYSPYGAFLEQIDQF
DAEFFGIVPREAVAIDPQQRLLLETTWQALESAGQNPQKLRNTQTGVFVGAMTQDYAQLS
YAPEAINAYTGSGTSLSVAAGRISYVLGLQGPSMTVDTACSSSLVAVHLACNALRNGECD
IALAGGVNIILTPVISLIESRAHMLAPDGRCKTFDASANGMVRGEGCGMIVLKRLSQAVK
SGDHILAKVHSTAVNHDGSSSGLTVPNGDAQEKLLHQALKAAKLNPEQIDFIEAHGTGTA
LGDPIELESMAAVFGKRLQNRPLIIGSVKTNLGHLEGAAGIAGLIKTVLALQHHKIPPHL
HFQQPNPRFDWSSQIFEVPVHGKNWHPSQRERIAGVSSFGFSGTNAHIIVGEIASNSPQP
SEQKFYLLPLSARSQKSLKELAKNYQYALNESVNFADTCFTASTGRAIFRHRLCVLADSN
TTAEKALADFQKGEDSDNLITPITSETQTKVVFLFSGQGSQYSGMGQTLYNQEPVFKNTL
ELCDNILQPILGKSLLGLIFQLQNSEQLEQTQITQPALFSLEYALAKLWQSWGIQPAALL
GHSIGEYVAACLAGVFSLEDALQLVVQRGRLMGELPHNGAMAAIYADYQTVADHLTPYGN
QVNIAADNGAINVISGLSEIVEQLEKSFMEQGYKTRRLAVSHAFHSPLMEPILDDFAKML
QQVSFHEPSLNIISNVTGKPIGKEIATADYWLRHLRNTVHFGQGFKLLIDSGYRHFLELG
PKPVLLGMARLNSQSKEILWLPSIVPGQDEQAQMYRSLATLFVNGYSIEWTEVFKQVQRV
LLPTYPFQRERYWLSNPQFFVSDIKTKLHPFIHEVKKLATGEIIVEGEISSEYPDYLEGH
KVFGKILFPATGFIETILVASRQIFTDDLAIIQNVAIHQGLVLSQSQTKLQLIFKSKKSS
GYSFEIFSSDSAKDSWVLHVTGEINVNERHEKLTHRRRDAKDERRESVDLARFYEFYEQM
GISYGQRFQAIQELRYFSSSSQAKISIDRSLVDRRYCLHPVLLDACLQSIGAAFPEIHGQ
ELHLPYGFSSLELFGNPGTQAWTHIEIKSQIDGEICVDVEVYDEQEQLCARFTDLTARRI
NPAVLQRLWQETENNCFYQVQWQKLDSVSTITGNSQHSWLVFVRPSTALYQSINLLQKAG
ERVITVELSDDYKRHSLESFVINPSRKSDFQRLYQEAYPSGEFPTGVIFAWETVPNEASA
DTVYKSCNAVLYLIQTITSNMKKLPDLWLVTRGANRVLSETYLQPEQSPLWGLGAVINHE
YPQIRCVCLDLPAIVESHEAEFLFNEFHTSGSESRLALRRGNRYGARLVSATIPAAQKQQ
LVSKEGAYLITGGLGKLGLLMAQWLSQMGSSHLVLCSRHVKSQPEAIASLTKNGTQITIV
NADITSAADTEQLFSRFGADLPPLRGVIHAAAVLDDGLLTNQNWEKYQNVMRPKVEGTLL
LDRYTRNLSLDFFIAFSSAAVILGSPGQSSYAAANAFMDALIQQRQSLGLPGISIKWGAW

Fig. 23A

DTGNKIEKQRFANWGIHSMPSDTAIKYLSDLILSDVDQGIILDIDWSTFNQAFNINQPFF
AEVITTKADSKEAKLLERLKSVSIDERAENLSQGIEQILREVTGLSASSVIPHHTSFLEL
GLNSLMVLEFKNRLQSNLACTLPTSIIFDYPNIASLNIYLQKEVLADSVDFEIKSNESSE
IVNPYESLNEDELAILLNQKLAELEEYGD

Fig. 23B

Anabaena mcyD Length: 11610 SEQ ID NO 70

```
1 ATGGAGTTGG AGATAAGTAA CGATAAGTCT GTCCAAGCTA GGGTGTTTAA
51 GGCGCTAAAT GAAGCGAAGG AGAGATTAAA AGCTGTTGAA GCTCAACAAA
101 GCGAACCGAT TGCGGTTGTG GGTTTGGGTT GTCGTTTTGG TAAAAATATT
151 TCCACTCCTG AAGCTTTTTG GTCTTTTCTC AAAGCTGGTG GTAATACGCT
201 CACAGATGTA CCTAGCGATC GCTGGAATGT GTCTGATTTT TATGACTCAG
251 ATCGTAGTAA ACCCGGCAAA ATCTATGTAT CTCAAGGCGG TTTTCTCAAT
301 GACGTGTCGA TGTTTGACGC GCACTTTTTT GGGATTGCAC CGAGAGAAGC
351 TTTGCATATC GATCCTCAGC AACGGTTACT TTTAGAAGTT GGCTATGAGG
401 CGTTAGAAGA TGCTGGTATG GCTGCGCCTA CTGCCAGAAT CGGCAAAACC
451 GGTGTTTTTA TTGGCATCAC AAACAATGAT TATGCTCGTC TGATTGCACC
501 GGGAGGAGAT TATAGCACAA TAGGGGCTTA TCATATCTCT GGCAACCATA
551 TAAATGCTGC TGCTGGACGC ATTTCTTATC TTTTGAATCT CAATGGCCCT
601 AGTTTGGCTA TCGATACGGC TTGTTCTTCT TCACTGGTAG CAGTGCATTT
651 AGCTTGCCGT TCTTTGCGCT CCCAGGAATG TCGGCAAGCT TTGGTGGGGG
701 GAGTAAACCT AATTCTCACC CCGGAAGTGC CGATCGCCTT GTGCAGAAAT
751 CAGATGTTAG CTGCTGACGG TAGGTGCAAA ACCTTTGACG AAAGTGCTGA
801 CGGATTTGGG ATTGGTGAAG GTTGCGGAGT TGTTGTTTTA AAAAGGTTGA
851 GTGAAGCCTT AGCAGATAGT GATCGCATTT GGGCGTTAAT TCGAGGAACA
901 TCTGTCAATA ACGATGGTGC CAGTGGCGGT TTTACTGTTC CTAATGGACC
951 TATGCAGACA GATTTAATTC GTGATGCTTT AGCTGATGCT CGTCTAAGTG
1001 CCAATGCTAT TGATTATGTC GAGGCTCACG GTACGGGTAC ATCTTTAGGA
1051 GATCCGATTG AAATTAAAGC GATCGCTGAT GCTTTATGTG TCAATCGCTC
1101 TCCATCCCAA CCTTTGTTAA TCGGTTCTGT CAAAACAAAT CTTGGTCATC
1151 TAGCGGCGGC GGCGGGGATT TCTGGCTTGA TCAAGACTAT ATTATCTATC
1201 TATCATGGCG AAATTCCGGC ACATCTCAAT TTCCAAAAAC CAAATCCTCA
1251 CATTAATTGG GATAGCATAC CTCTGCACGT GGTGACACAG ACTCGCCCCT
1301 GGACTCAAGG CAATGGCAAG ATAAAAGCGG CTGGAGTCAG TTCTTTTGGA
1351 GCATCGGGAA CAAATGCTCA TGCAATATTG AGTGAACCTC CGCAAGTACA
1401 ATCCTCAACT AAAACAACAT TTCCTGTACT GGTAACTCTG TCTGCTAAAG
1451 ATGAAGAACG ATTGCGTCTG TTGGCTGCTA ACTTTGTTAA TTATATTGAC
1501 CAAAAACCGG AAGTTAACTT AACTGATATC GCCTTTTCTG TAAATACAGG
1551 TCGTTTTCAT CATAAACAGC GATTAACTTT ATTAGTTTCC AATACTGCTA
1601 GTTTTAAAGA ACAACTTCGA GCTTTTGTTA ATGGTAATGA CCGAGAAGCT
1651 AATATCATTT ACGGTGAGGT ACGAGTAAAT ACTAAAATCG CTTTTGTTTG
1701 CAATGATGCA AGAACTATAT TTCCGGCGAT GGGGAAAGAG CTTTTTGTAA
1751 ATGAGCCTGT ATTTCGTGAT GCTTTGATGC AATGCGATCG CCTATTTCAA
1801 AACTATTTAA ATTTCTCTAT AGAAGACTTA CTTGATAACA ATCAAGATAT
1851 AAACGATCCT TTATATGTTC ACCCAACATT ATTTGCGCTG CAATATGCTC
1901 TTTGTGAACT CTGGAAATCT TGGGGAATCC GCCCTTCTGC AATTTTAGGA
1951 AATGGTTTAG GCGAATATAT TGCTGCTCAA CAAGCTGGAA TTTTTTCGAT
2001 TGGAGATGCA GTTAAATTAG TGGCAACTCG CGTTCATTTA TTTAGCGAGC
2051 AAAGTGGAGA AAAAGAATTT TTAGCTTTTC AAAAAGTAGC TGAAGAAGTT
2101 TCTTATCGTA ATCCACAACT CACTTTAATT TCTAATAATG GTGAGGTTGC
2151 TAATAATTAT ATTAAAACTG CTGCTTACTG GTTGCAACAG CTACATCACC
2201 TAGACAATAT TGAAAAAGGG ATTTTTACCT TACATCAGAT GGGCTATCAA
2251 ATATTTCTAG AGATTGGAAA TCAGCCCATA AATAATATCG TTCATAGTAA
2301 ATTACCAGCA AAAGAAATCC TTTGCTTATG TAGTTTAGCA AAAGATTGCA
2351 GTAATTATCA GCAAATCCAA ACAGTCATAG CTGAACTATA TGTAAGAGGA
2401 GTAAATATTA ATTGGCAAGA ATTCCATCGC TCTTGCTCAG GACAAAAAAT
2451 TAGTTTGCCC AGTTCTCCTT TTATTCGCAA ACGCTACTGG GTAGATGCTA
2501 TTAAATCAAA TCGCTCTGGA CACCCTTTAT TAGGCGATCG CTTCCCCTCT
2551 CCCCTCTCTT CTATTCAATA TCGAGCTAGT ATTAGTCAGA ACAACCCAGA
2601 ATTTTTAACA GAACATCAAG TTTTTGATAA ACCGATATTT CCTGGTGCAG
2651 CTTTTATTGA AATGGCGTTA GCAGCGGCAC AAAGCCAAGC TGTAATTTTA
2701 GAAAATATCG AATTTCAAAA AGCACTATTG CTGCAAGAAC AAGAGGATAC
2751 ATTACAGTTA ATTATTGAAC AAAATTGTTT TCAAATATAT CAGCAAAGTA
2801 ATAATAATTG GGATGTACTT GTTACCGGGG GAATTGATAG TTTAAATATA
2851 CTAGATGCAA ATCGCCAAAA TCTAGAACAA ATAGCTGCTA ACTGTCCAGA
2901 GCAGATGGAA ATTAATTCAT TTTATGAAAG ATATCAAAAA TCAGGAGTTA
2951 ATTACGGCAG TAGCTTCCAG ATCATTAATA AGCTTCAACG TGGCAAAAAT
3001 AGTGCATTTG CTAAAATCAA GTTAACACAA ATTCAGTGTC TAGAAGCCGA
```

Fig. 24A

```
3051 AAAATATAAG TTTCATCCAG TAATGTTGGA TGCTTGTTTT CAGGCGATCG
3101 CTGCAATTTT ATTTCAGCAA GAATCATCTG TAACTTATGT CCCAGTCCGC
3151 ATTGCTAAAT TTCAATTCTT CAAGTCCCCA GGTGCTAGTG TCATCAGTGC
3201 AGCTCGACTA ACTAAAAATT CTAATAACTA TATTATCTCT AATATTGATA
3251 TTTATTCAGA ACAAGGTGAG TTACTAGTAG CAATTACAGG ATTTGAACTT
3301 AAATCAGTTC AAAGTCAAGA AATTATCCGC CATCAAATCC AACCACAATC
3351 TTACATGGAG GAGTGGATTA CTTTAGCATC ATTGTTACCT GATGGTAGAG
3401 ATTATTTACT AAACCCTGAT GCCATTAAAA CTCTAGTTCA ACCCCAGCAA
3451 TTAGAACAGC AATTAGCTGA AAAATTGTAC CAATATGAAC GCTTGCTTGA
3501 AGAAATGGAA ACTTTAAGCG TAAGTTATAT ATGGGAAGGT TTAAAAGAGC
3551 TTAACTGGCA ACCCCAATTA GGTCAAATTT ATCCAGAAGA ACAGATTGCT
3601 ACTCAAGGTG GAGTTGTTGA TTTTTATCGT CCTTTATTGT CGCGGTGTTT
3651 AGCAATTCTA GCTGAAGAAG GAATAATTAC TCAGCAAAAA GATGGATGGT
3701 TATTAGCAAA AGAACCTGTT ATATCTTCAT CTCAATTACC AATTCAACAA
3751 CTACGTAGGG AATTTCCTGA TTACCTTGCA GAAATTAATT TGATAGAACG
3801 TTGTGGTTCG GCTTTAGCTG CGGTAATGCG CCGCCAAATT GAGCCATTAG
3851 AATTACTATT TCCCCAAGGC GATTTGAATG CGATCGCATC TGTATATTCT
3901 GATGCGGCTG GTGCTAAGTT GATGAATGAA CTAGTCGCAG CCACAATTAA
3951 AACTGTAGTA GCAAATTTAC CAACAAATCG ACAGTTACGC ATCCTCGAAA
4001 TTGGAGGGGG AACAGGCTCA AGTACCGCAG CAATTTTACC ACATTTACCT
4051 CCAGAACAAA TTGAATACAC ATTTACCGAT ATTTCTTCTA GCTTTTTAAC
4101 CAGAGCCAAA GAAAATTTCA GTAATTATCC CTTCATTAAA TATCAAACTT
4151 TAGATATAGA AAAAGCCCCC ATTTCTCAAG GATTTCTACC TAGTTATTTT
4201 GATATTATTA TTGCTGCTAA CGTCCTTCAT GCCACAGCAG ATATTAATGA
4251 AACCCTTAAT AATGTCCGTT CTTTACTAGC TCCTAATGCC ATACTCATAC
4301 TATTAGAAAG TACCGGAGCG CGTCCTTGGG TAGATTTAAC CTTTGGGTTA
4351 ACAGAAGGAT GGTGGTTATG CAGTCAAGAT CCGCACCGCA ATGGTTATCC
4401 ATTAGTAGAT ACAGAACATT GGCAAAACCT ACTCGCCAAG CATCAATTTA
4451 CAGAAATTAA CATCATCGAA CCAACCAATC CAAAGACACG GAACCTATTA
4501 CAACAAAGCG TAATCATCGC CAAAAGCAGC CTTCCTTCTC TATGCACCTC
4551 TGTGTCTTGG CGCGAAATAA TCTTTGCTGA CACTAACGGA ATTGCTCGCA
4601 GTTTAATCAC CCCATTCCAA CAGCGAGGAA TTACTTGTTC CCTCATCTCC
4651 CCTCAAGACA TTAACCCCGA TAACCCCGAC GATTACCTAT CACTATTACA
4701 AAACCTGATA ACACCAGAAA CCCGTGAAAT CATCTACCTC TGGTCATTAC
4751 AAGAAATAGA AGGCGAAATT TATCAAGCTG TAGAAATTCA TTGTAGACGT
4801 TTTCTATTTC TACTTCAGGC ATTATTACAA CAAGAAAACC CCCCCGCGCT
4851 CATTCTTGTT ACTCAAGGGT CTGTACCAGC AAAAGAAATC ACAACCTTAA
4901 CTTCTCCAGC CCAATCTTCC TTATTAGGAA TGGCTTTAAG CTTAGTTCTA
4951 GAGCATCCCG AACTAAATTT TAGAGCGATT GATTTAGATC CTCATGCACA
5001 AGATTTAGGA GAAAAATTAT TTAGAGAAAT CCACAACAAC ACCCAAGAAA
5051 ACCGCGTCGC TTTGCGGGGC GAACAACGCT TCTGTCCGCG TTTGGTAGAG
5101 CGCAAATTAG CAGATGGAAA CATCAATTTC CGTCAAGATG GCTTTTATCT
5151 AATTTCTGGT GGTACAGGCG GATTAGGTTT AGCCACTGCC CGATGGATGA
5201 TAGAACATGG CGCGTGCCAT TTAGTATTGT GCAGTCGCAG TGGTGCAAAA
5251 GCTCTCAATC CTGAAATTCT AGCAAGTCTG CAATCCATTA ACGAAGACAT
5301 CCAAATTAAA GATGTTGATG TCACAGATGC CGAGAAACTG CACGCCCTAC
5351 TAGAAGAATG CCGATCGCAA TATCCGTTGC GAGGAATTTT CCACATCGCT
5401 GGCACTTTAG ACGATACAAC ATTACTGCGA CTGACACCAG AACGTTTCAA
5451 CTATGTTCTC GCCCCCAAGG TGAAAGGAAC TTGGCTACTG CATCAACTAA
5501 CCCTGAACGA TACCTTAGAT TTCTTTGTTT GTTATACCTC TGCTGTGTCG
5551 CTCATTGGTT CGGCGGGACA AGCTAATGCT GCGGCTGCGA ATGCCTTTGA
5601 AGATGCCTTT ACCTATTATC GTCATGCACA CAATTTACCA GCTACAGTCA
5651 TCAACTGGGG GCCGTGGAGC GAAATCGGTG CAGCAGTCGA TCGCAATGTC
5701 TTAGAAAGAT TAGCAGCCAA AGGTTATGAT GCGATCGCAC CAGATTTAGC
5751 ACTCAATACT CTAGAAAAAA TTCTGTTTAA TCAAATTGTA CGTGCTGGTG
5801 TAATTGCTAT TGATTGGCAA CGTTTTCCAT ACATTAATCA AAGCTTTTAT
5851 CAAAACTTCT TACCCCAAAT CAAACCTAAA TCTCAAACTG CATCCAATCT
5901 TTTAGAACAG TGGCAAATCA TACCTGTAAA ACAGCGTCGT GATTTACTTA
5951 TCCGCCAAAT TAGTTTGCGA GTTTGCACTG TATTAGGACT ATCAACTCAT
6001 GAAGTATCTC CCCAACAAGG CTTTTTTGAC ATGGGGATGG ATTCTCTCAC
6051 TTCCACAGAA CTTCGCAATC TTCTGCAAAC AGATTTTAAT TGCTCTCTAC
6101 CCACAACAAT CGCTTTTCGG TTTCCCAATG TGGAAACTTT AGCAGATTAT
6151 CTCCTGCGGG AAATTTTAGT TACCTCTGAG GTGCAAACTC CTGTACAACA
```

Fig. 24B

```
6201 ACTAATACAA GAAGTTCCCC AAATCCAAAT TGAAAAATAT ACCCCCGAAA
6251 AACCGCAACA GGAAGAAGAT CCAATTGTGA TTGTAGGGAT GGCTTGTCGA
6301 TTTCCTGGCG GTGCGAATGA TTTGGAATCA TTCTGGCAAC TTTTAGAACA
6351 GGGGAAAGAT GCAGTCAGAG AAATTCCCAG CGATCGCTGG GATATGCAAG
6401 CATGGTATCA TCCCGATCCA GATACACCAG GAAAAATCTA TTCTCCTTAC
6451 GGAGCATTTC TAGAGCAAAT AGATCAATTT GATGCCGAAT TCTTTGGCAT
6501 TGTCCCCCGC GAAGCAGTCG CCATCGATCC TCAGCAACGC CTACTATTAG
6551 AAACAACTTG GCAAGCCTTA GAATCTGCTG GGCAAAACCC CCAAAAGTTA
6601 CGCAACACTC AAACAGGGGT TTTTGTGGGG GCGATGACTC AAGACTATGC
6651 CCAGTTGAGT TACGCACCAG AAGCAATTAA CGCCTATACT GGTTCTGGCA
6701 CTTCTCTGAG TGTAGCTGCC GGTCGTATTT CGTATGTACT GGGCTTACAG
6751 GGACCATCAA TGACAGTTGA TACTGCTTGT TCTTCTTCTT TAGTAGCAGT
6801 ACATCTAGCT TGCAATGCTT TGCGTAATGG TGAGTGTGAC ATTGCCTTAG
6851 CAGGTGGTGT GAATATAATT TTAACTCCGG TAATTTCCTT AATTGAAAGT
6901 CGCGCCCATA TGCTTGCACC TGATGGACGT TGCAAGACTT TTGATGCGAG
6951 TGCTAATGGT ATGGTACGGG GCGAAGGCTG CGGGATGATT GTTCTCAAGC
7001 GGTTGAGTCA AGCCGTTAAA AGTGGCGATC ACATTTTAGC TAAAGTTCAT
7051 AGTACAGCCG TAAATCACGA TGGATCAAGT AGTGGTTTGA CTGTACCTAA
7101 TGGTGATGCT CAAGAAAAAT TATTACATCA AGCATTAAAG GCGGCGAAAC
7151 TCAACCCAGA GCAAATAGAT TTTATTGAAG CTCATGGTAC TGGGACAGCT
7201 TTAGGCGATC CGATTGAATT AGAAAGTATG GCTGCGGTAT TTGGCAAACG
7251 TCTCCAAAAT CGACCTTTAA TTATTGGTTC TGTCAAAACT AATTTAGGAC
7301 ATTTAGAAGG AGCAGCCGGA ATTGCTGGGT TAATTAAAAC TGTTTTAGCC
7351 CTACAACATC ACAAAATTCC TCCCCATCTT CACTTTCAAC AACCCAACCC
7401 CCGTTTTGAT TGGAGTTCTC AGATTTTTGA AGTTCCAGTA CATGGAAAAA
7451 ACTGGCATCC TAGCCAACGA GAACGCATTG CTGGAGTAAG TTCTTTTGGA
7501 TTTAGTGGTA CTAATGCTCA TATTATTGTT GGAGAAATTG CATCTAATTC
7551 TCCACAGCCA TCTGAGCAGA AATTTTACCT CCTGCCGCTT TCGGCTCGTT
7601 CTCAAAAATC TCTCAAAGAA TTAGCAAAAA ATTATCAATA CGCTTTAAAT
7651 GAATCTGTGA ATTTCGCAGA TACTTGTTTT ACTGCCAGTA CAGGAAGGGC
7701 TATTTTCCGG CATCGATTGT GTGTCTTGGC TGACTCAAAT ACTACAGCCG
7751 AAAAAGCACT TGCTGATTTC CAAAAAGGTG AAGATTCTGA TAATTTAATT
7801 ACTCCAATTA CATCAGAAAC TCAAACAAAA GTAGTTTTCC TATTTTCAGG
7851 ACAAGGTTCT CAATATTCAG GGATGGGACA AACTCTTTAC AACCAAGAAC
7901 CCGTCTTTAA AAATACTCTG GAACTTTGTG ACAACATTCT GCAACCTATT
7951 TTAGGAAAGT CGCTCTTAGG TTTAATTTTT CAATTGCAAA ATAGTGAACA
8001 GCTAGAACAG ACTCAAATCA CACAACCAGC CCTGTTTTCA CTGGAATACG
8051 CACTTGCGAA ATTATGGCAA TCTTGGGGAA TACAACCTGC TGCCCTTTTA
8101 GGACACAGTA TTGGTGAATA TGTCGCCGCT TGTTTAGCAG GAGTCTTTTC
8151 TTTAGAAGAT GCTTTGCAGT TAGTTGTTCA ACGCGGTCGC TTGATGGGAG
8201 AACTACCGCA TAATGGAGCA ATGGCAGCAA TTTATGCTGA TTATCAAACA
8251 GTAGCCGATC ATCTGACTCC CTATGGAAAT CAAGTAAATA TTGCCGCAGA
8301 TAATGGTGCA ATTAATGTTA TTTCTGGCTT GAGTGAAATT GTTGAACAAC
8351 TAGAGAAATC ATTTATGGAG CAAGGTTATA AAACTAGACG TTTGGCAGTA
8401 TCTCATGCTT TTCATTCTCC TTTAATGGAG CCAATCCTTG ATGATTTTGC
8451 TAAGATGTTA CAGCAAGTTT CTTTCCATGA ACCAAGCCTG AATATTATTA
8501 GTAATGTCAC AGGTAAACCA ATTGGCAAAG AAATTGCTAC TGCCGATTAT
8551 TGGCTGCGCC ATCTTCGCAA TACTGTACAT TTTGGTCAAG GATTTAAATT
8601 GTTAATTGAC AGTGGTTATC GCCATTTTTT GGAGTTAGGA CCAAAGCCTG
8651 TGTTATTAGG TATGGCGCGG CTAAATTCTC AAAGCAAAGA AATATTATGG
8701 CTACCCAGCA TTGTACCTGG ACAAGATGAA CAAGCCCAGA TGTACAGAAG
8751 TTTAGCAACA CTTTTTGTTA ATGGTTATTC TATAGAATGG ACAGAGGTAT
8801 TTAAGCAAGT CCAGCGGGTT CTTTTACCTA CTTATCCGTT TCAACGAGAA
8851 CGTTATTGGT TATCAAATCC ACAATTTTTT GTAAGTGATA TTAAGACTAA
8901 ACTGCATCCT TTTATTCATG AGGTTAAAAA GCTAGCTACA GGAGAAATTA
8951 TTGTTGAAGG AGAGATAAGT TCTGAGTATC CAGATTATCT GGAAGGACAC
9001 AAGGTATTTG GAAAAATCTT GTTTCCGGCT ACGGGTTTCA TTGAAACGAT
9051 TTTAGTAGCT AGTCGGCAAA TATTTACTGA TGATTTGGCG ATTATTCAAA
9101 ATGTTGCTAT TCATCAGGGG TTAGTTTTAT CTCAATCTCA AACTAAATTG
9151 CAACTGATTT TTAAGTCTAA GAAGTCGTCT GGCTATAGTT TTGAGATATT
9201 TAGTAGTGAT TCAGCAAAAG ATAGTTGGGT ACTTCATGTT ACGGGTGAGA
9251 TTAACGTTAA TGAAAGACAT GAAAAGCTCA CACACAGACG CAGAGACGCA
9301 AAGGATGAAA GAAGGGAATC TGTAGATCTA GCCAGATTTT ATGAGTTTTA
```

Fig. 24C

```
 9351 TGAGCAGATG GGAATTAGCT ATGGTCAGCG TTTTCAGGCT ATTCAAGAAC
 9401 TCAGGTATTT TAGCAGTAGT TCTCAGGCTA AAATTAGTAT CGATCGCAGT
 9451 TTGGTAGATC GGCGATATTG TTTGCATCCT GTGCTTTTAG ATGCTTGTTT
 9501 ACAAAGTATT GGCGCGGCTT TCCCAGAAAT TCACGGTCAG GAATTGCATT
 9551 TACCTTATGG TTTCTCTTCT CTAGAATTAT TTGGCAATCC TGGTACTCAA
 9601 GCTTGGACGC ACATTGAGAT TAAATCTCAG ATTGATGGTG AAATCTGTGT
 9651 AGATGTAGAG GTTTATGATG AGCAAGAACA GCTTTGCGCT CGGTTTACAG
 9701 ACTTAACAGC ACGGCGGATT AATCCAGCAG TTTTACAACG TTTATGGCAA
 9751 GAAACTGAAA ATAATTGTTT TTACCAGGTG CAATGGCAAA AGCTTGATTC
 9801 AGTTTCAACT ATTACAGGAA ACTCTCAACA TTCATGGTTA GTGTTTGTAC
 9851 GTCCTAGCAC AGCTTTGTAT CAGTCGATTA ATTTATTGCA AAAAGCTGGA
 9901 GAAAGAGTTA TTACTGTAGA ACTTAGTGAT GATTATAAAC GTCATTCTCT
 9951 AGAAAGTTTT GTCATCAATC CATCTCGGAA ATCTGATTTT CAACGTTTAT
10001 ATCAAGAAGC ATATCCATCG GGTGAATTTC CTACGGGTGT GATTTTTGCT
10051 TGGGAAACTG TACCTAATGA AGCAAGTGCA GATACAGTTT ACAAGAGTTG
10101 TAATGCTGTA TTGTATTTAA TTCAAACTAT TACTAGTAAT ATGAAAAAAT
10151 TACCTGATTT ATGGTTAGTA ACTCGTGGTG CGAATCGAGT TTTATCTGAA
10201 ACTTATTTGC AACCGGAACA ATCTCCGCTT TGGGGATTAG GAGCAGTAAT
10251 TAATCATGAA TATCCCCAAA TTCGTTGTGT ATGTTTAGAT TTACCAGCAA
10301 TTGTAGAATC TCATGAAGCT GAATTTTTGT TCAACGAATT TCATACATCA
10351 GGCAGTGAGT CACGCCTAGC TTTACGCCGA GGAAACCGTT ATGGAGCAAG
10401 GTTAGTATCT GCAACAATTC CCGCAGCCCA AAAACAGCAA TTAGTTAGCA
10451 AAGAGGGAGC TTATTTAATT ACAGGTGGTT TGGGTAAACT TGGCTTGCTC
10501 ATGGCTCAGT GGTTATCACA AATGGGATCT AGCCATTTGG TTTTGTGTAG
10551 TCGCCATGTC AAATCTCAAC CAGAAGCGAT CGCCTCATTA ACTAAAAATG
10601 GTACACAGAT TACAATTGTC AATGCTGATA TTACCTCAGC CGCAGATACG
10651 GAACAGCTAT TCTCCCGTTT TGGCGCGGAT TTACCCCCCT TACGCGGAGT
10701 CATTCACGCG GCAGCTGTCC TTGATGATGG GCTTTTAACT AATCAAAATT
10751 GGGAAAAATA TCAAAATGTT ATGCGTCCCA AAGTGGAAGG CACTTTATTA
10801 TTAGACCGCT ACACCCGCAA TCTATCTTTA GACTTTTTCA TAGCCTTTTC
10851 TTCTGCTGCT GTAATTCTAG GTTCACCGGG ACAAAGTAGT TACGCTGCTG
10901 CTAATGCTTT TATGGATGCT TTAATACAGC AACGACAAAG TTTAGGATTA
10951 CCAGGTATTA GTATCAAGTG GGGCGCATGG GATACAGGGA ATAAAATTGA
11001 AAAACAACGT TTTGCTAATT GGGGTATTCA CAGTATGCCT TCTGATACCG
11051 CTATCAAATA TCTCAGTGAC TTAATTCTCA GCGACGTTGA TCAAGGAATT
11101 ATTCTTGATA TTGATTGGTC TACTTTTAAT CAAGCTTTTA ACATTAATCA
11151 ACCATTCTTT GCCGAAGTAA TTACAACTAA AGCAGATAGC AAAGAGGCAA
11201 AATTATTAGA ACGATTAAAG TCTGTATCTA TAGATGAGCG AGCAGAAAAT
11251 TTATCGCAAG GAATTGAGCA AATTTTACGC GAAGTTACCG GATTATCTGC
11301 TAGTAGTGTG ATTCCTCACC ACACCTCATT TTTAGAGTTG GGTTTAAACT
11351 CGTTAATGGT ACTAGAATTT AAAAATCGGC TGCAAAGTAA TTTAGCTTGT
11401 ACTTTACCTA CATCTATAAT CTTTGATTAT CCTAATATTG CGAGTCTCAA
11451 CATTTATTTG CAGAAAGAAG TGTTAGCAGA TTCTGTTGAT TTTGAGATAA
11501 AATCAAATGA ATCTAGTGAG ATTGTCAATC CTTATGAAAG TCTCAATGAA
11551 GATGAGTTGG CTATACTACT TAATCAAAAG TTAGCGGAGC TTGAAGAATA
11601 TGGTGATTAA
```

Fig. 24D

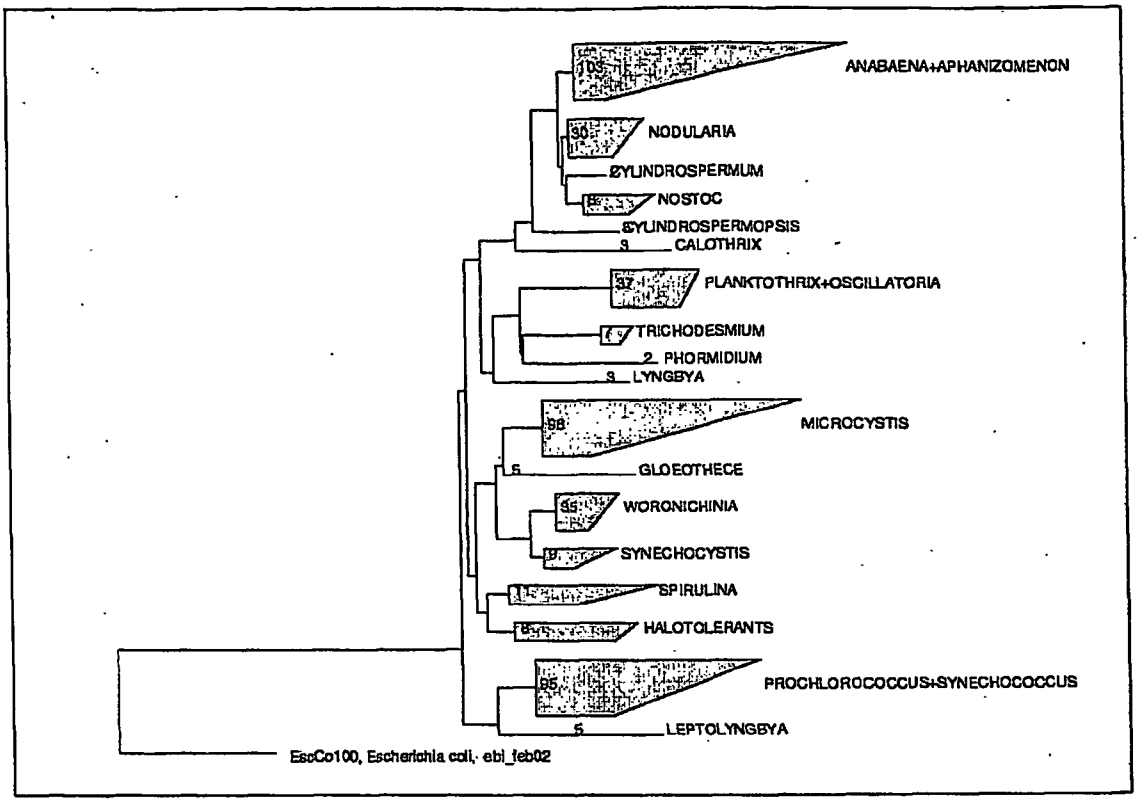
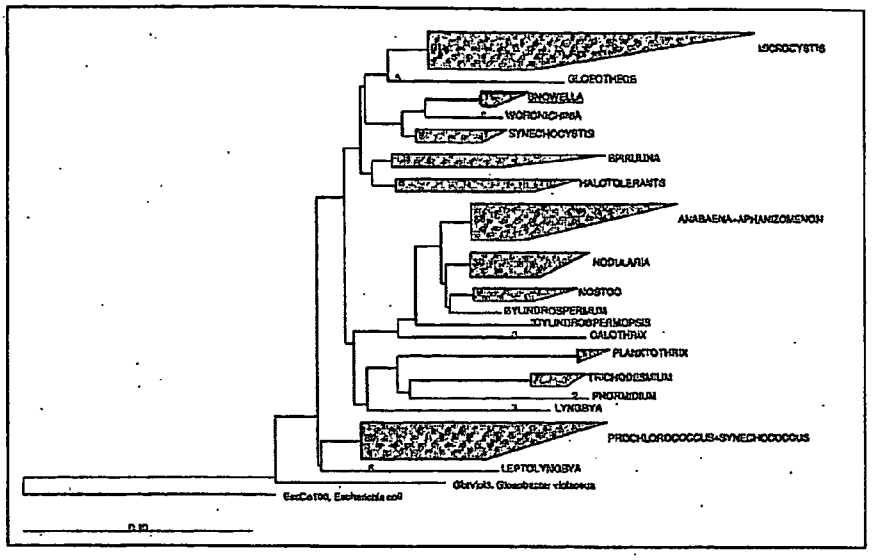

ANABAENA+APHANIZOMENON
NODULARIA
CYLINDROSPERMUM
NOSTOC
CYLINDROSPERMOPSIS
CALOTHRIX
PLANKTOTHRIX+OSCILLATORIA
TRICHODESMIUM
PHORMIDIUM
LYNGBYA
MICROCYSTIS
GLOEOTHECE
WORONICHINIA
SYNECHOCYSTIS
SPIRULINA
HALOTOLERANTS
PROCHLOROCOCCUS+SYNECHOCOCCUS
LEPTOLYNGBYA
EscCo100, Escherichia coli, ebi_feb02

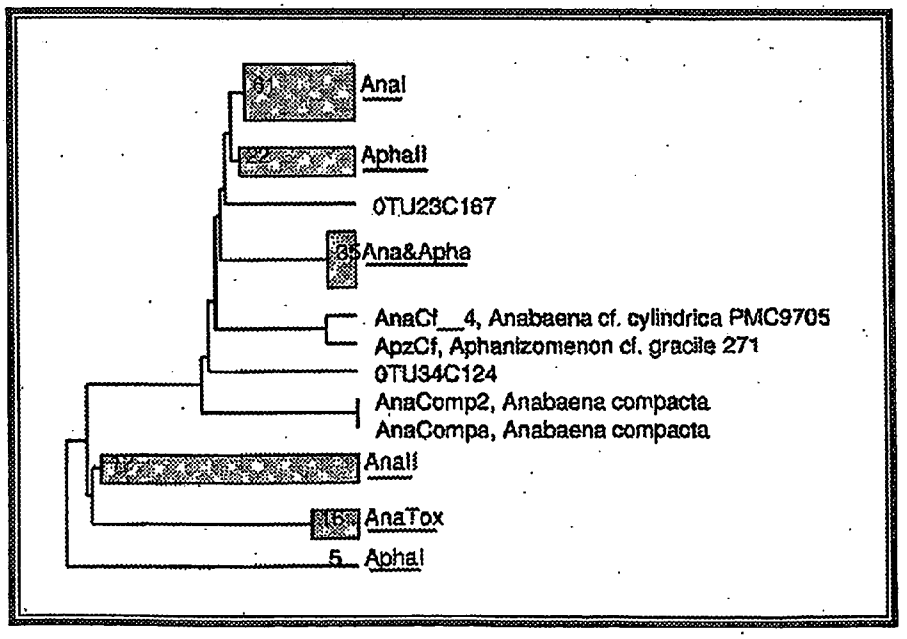
Fig. 25A

Fig. 25B

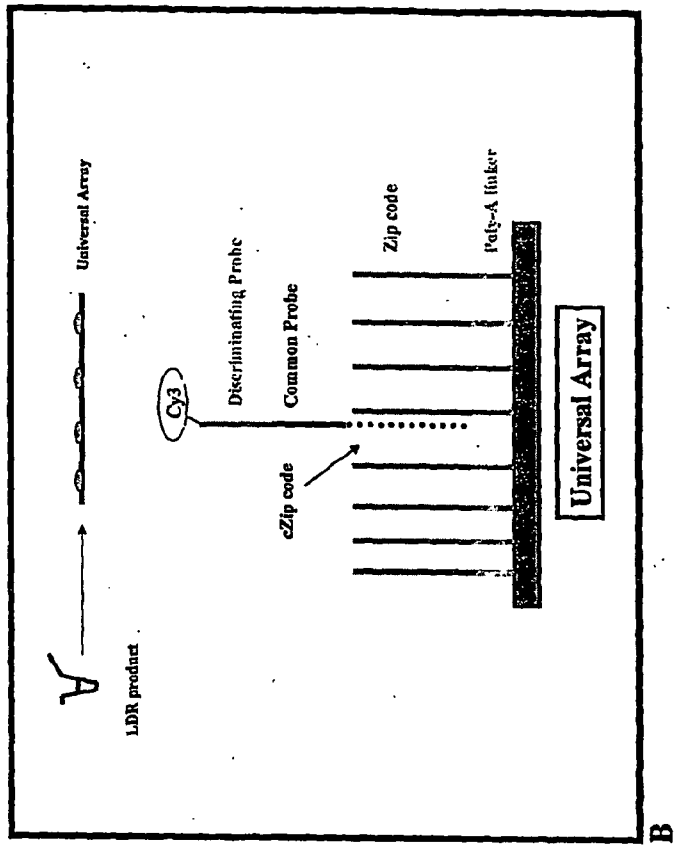


New cyanobacterial 16S rDNA sequences belonging to *Snowella* spp.

Fig. 25C



New Microarray with the subdivision of *Aphanizomenon-Anabaena* in 6 subclusters


Universal Array
Discriminating Probe
Common Probe
Zip code
Poly-A linker
Cy3
cZip code
Universal Array
LDR product
A
B

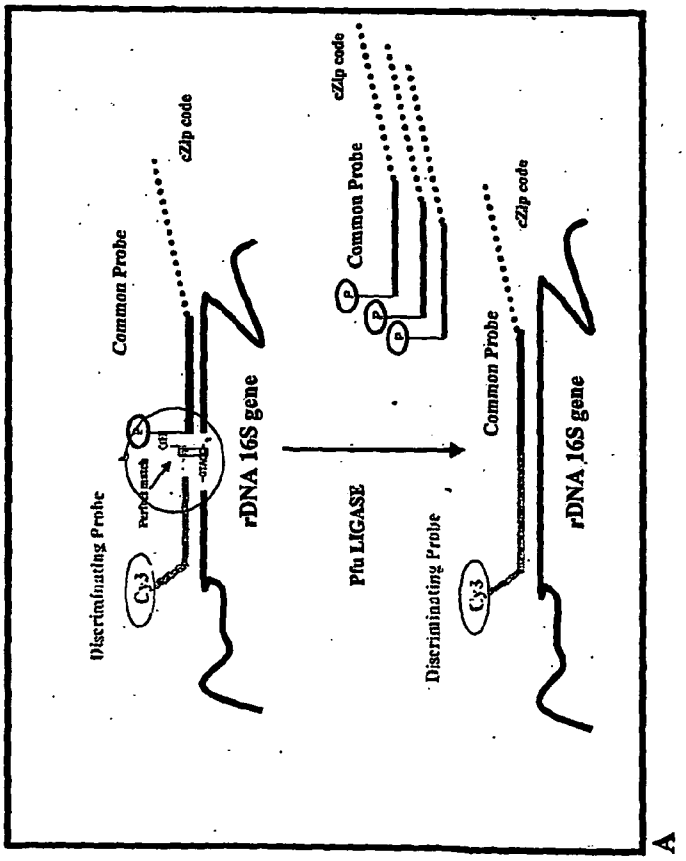
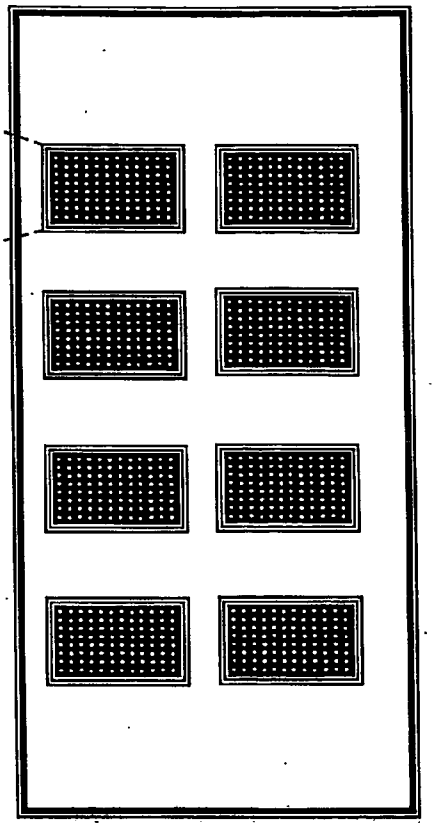

Discriminating Probe
Common Probe
cZip code
rDNA 16S gene
Cy3
Pfu LIGASE
A

Fig. 26

1	1	25B	25B	1B	1B	1B	1B	25B	25B	1	1
3B	3B	3B	3B	5B	5B	5B	5B	11B	11B	11B	11B
13B	13B	13B	13B	15B	15B	15B	15B	21B	21B	21B	21B
23B	23B	23B	23B	27	27	27	27	28	28	28	28
29	29	29	29	25B	25B	25B	25B	31	31	31	31
32	32	32	32	33	33	33	33	34	34	34	34
35	35	35	35	36	36	36	36	37	37	37	37
1	1	25B	25B	38	38	38	38	25B	25B	1	1



ZIP	1	Hybridization Control
ZIP	25B	Universal
ZIP	1B	Microcystis group
ZIP	3B	Synechococcus + Prochlorococcus group
ZIP	5B	Woronichinia group
ZIP	11B	Spirulina group
ZIP	13B	Halotolerants group
ZIP	15B	Prochlorococcus marinus group
ZIP	21B	Planktothrix + Oscillatoria group
ZIP	23B	Nodularia group
ZIP	27	Trichodesmium group
ZIP	28	Cylindrospermopsis group
ZIP	29	Cylindrospermum group
ZIP	31	Synechocystis group
ZIP	32	Nostoc group
ZIP	33	Phormidium group
ZIP	34	Lyngbya group
ZIP	35	Gloeothece group
ZIP	36	Calothrix group
ZIP	37	Leptolyngbya group
ZIP	38	Anabaena + Aphanizomenon group

Fig. 27A

Zip Number	Cyanobateria Group
1B	*Microcystis*
3B	*Synechococcus+Prochlorococcus*
5B	*Woronichinia*
11B	*Spirulina*
13B	*Halotolerans (Cyanothece)*
15B	*Prochlorococcus* marinus
21B	*Planktothrix*
23B	*Nodularia*
25B	Universal
2	*Oscillatoria/Planktothrix mcyE I*
10	*Snowella*
12	*Toxic Anabaena*
14	*Aphanizomenon Cluster I*
15	*Aphanizomenon Cluster II*
16	*Anabaena Cluster I*
17	*Anabaena Cluster II*
18	*Anabaena+Aphanizomenon SubCluster*
27	*Trichodesmium*
28	*Cylindrospermopsis*
29	*Cylindrospermum*
31	*Synechocystis*
32	*Nostoc*
33	Antartic *Phormidium*
34	Palau *Lyngbya*
35	*Gloeothece*
36	*Calothrix*
37	*Leptolyngbya*
38	*Anabaena+Aphanizomenon*
39	*Anabaena mcyE*
40	*Microcystis mcyE*
41	*Nodularia mcyE*
42	*Nostoc mcyE*
44	*Oscillatoria/Planktothrix mcyE II*

66	63	BLANK	1	2	3	4	5	1	2	3	4	5	BLANK	63	66
6	7	8	9	10	12	14	15	6	7	8	9	10	12	14	15
16	17	18	19	20	21	22	23	16	17	18	19	20	21	22	23
24	27	28	29	31	32	33	34	24	27	28	29	31	32	33	34
35	36	37	38	39	40	41	42	35	36	37	38	39	40	41	42
44	1B	3B	5B	11B	13B	15B	21B	44	1B	3B	5B	11B	13B	15B	21B
66	63	BLANK	1	2	3	23B	25B	23B	25B	1	2	3	BLANK	63	66
4	5	6	7	8	9	10	12	4	5	6	7	8	9	10	12
14	15	16	17	18	19	20	21	14	15	16	17	18	19	20	21
22	23	24	27	28	29	31	32	22	23	24	27	28	29	31	32
33	34	35	36	37	38	39	40	33	34	35	36	37	38	39	40
41	42	44	1B	3B	5B	11B	13B	41	42	44	1B	3B	5B	11B	13B
66	63	BLANK	66	15B	21B	23B	25B	15B	21B	23B	25B	66	BLANK	63	66

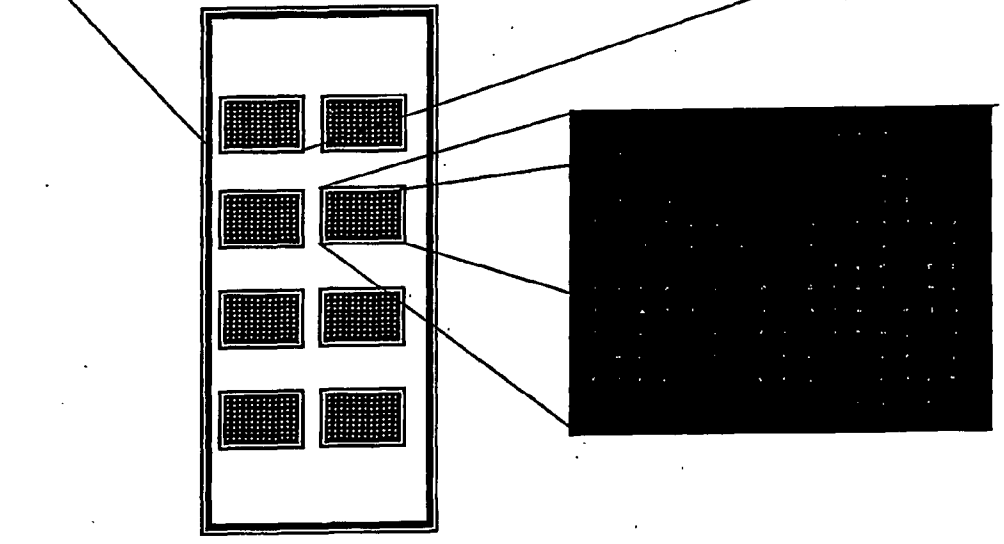


Fig. 27B

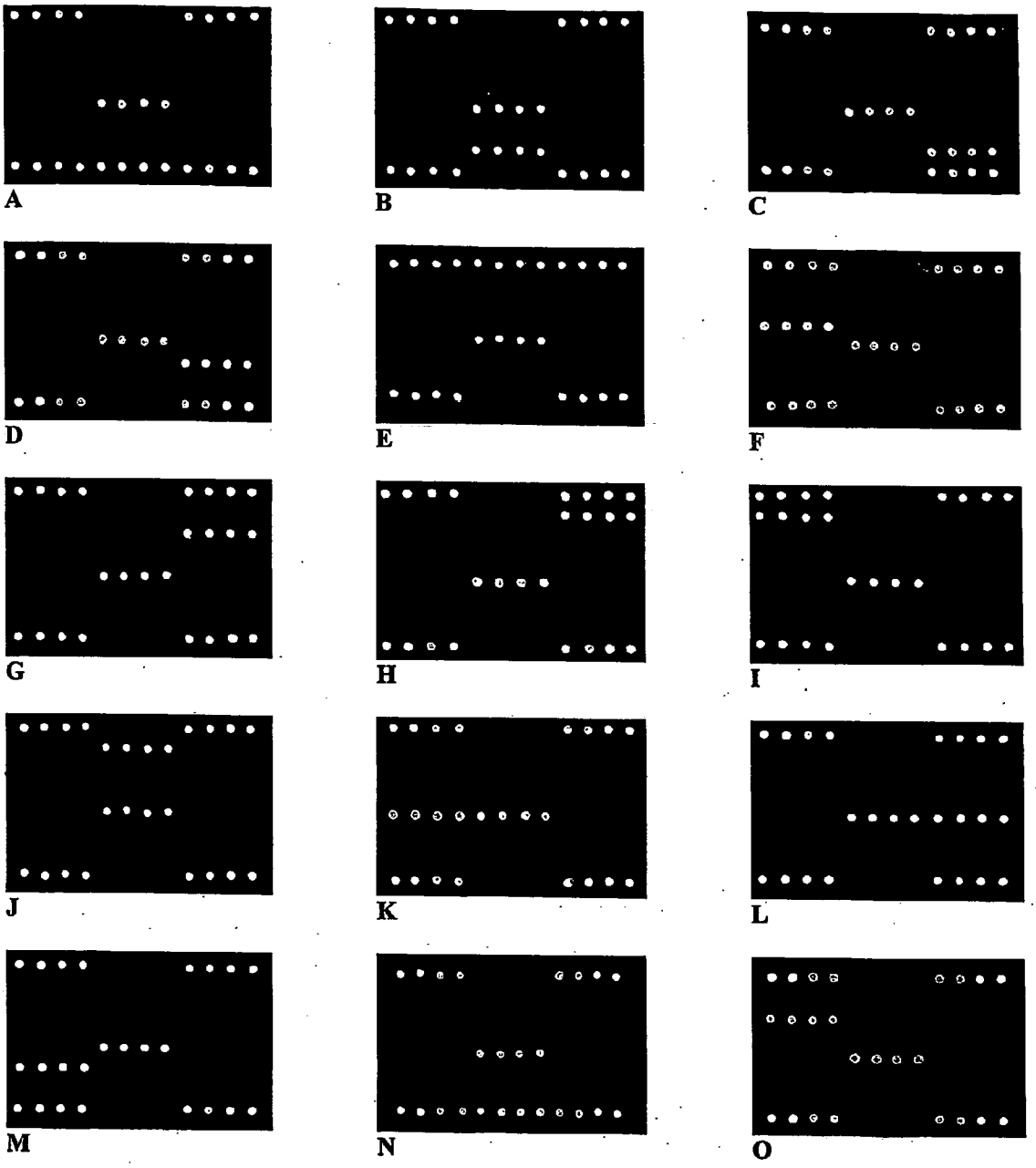
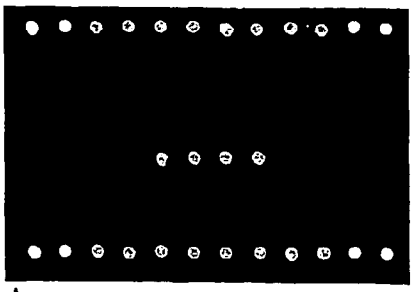
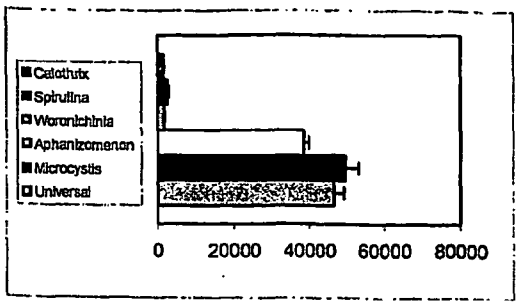
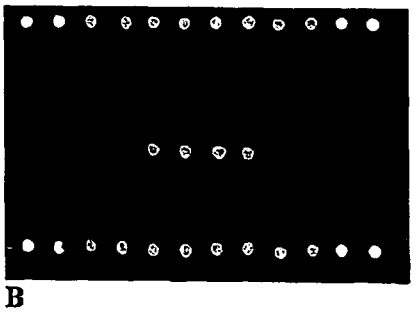


Fig. 28

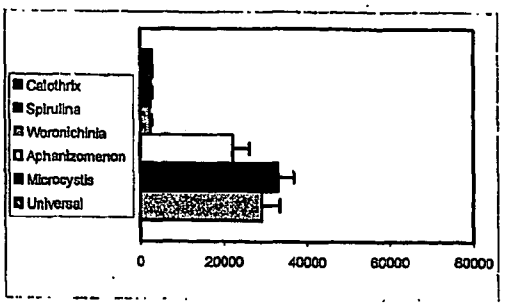


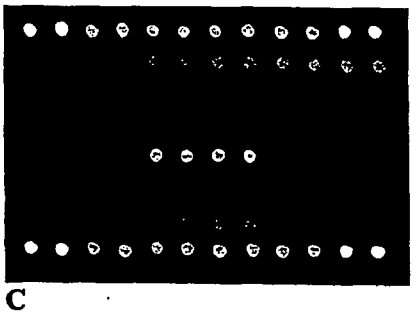
A


Calothrix
Spirulina
Woronichinia
Aphanizomenon
Microcystis
Universal
0
20000
40000
60000
80000

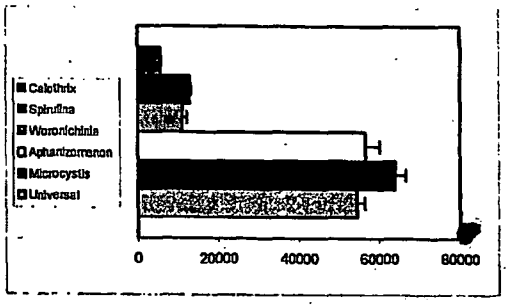


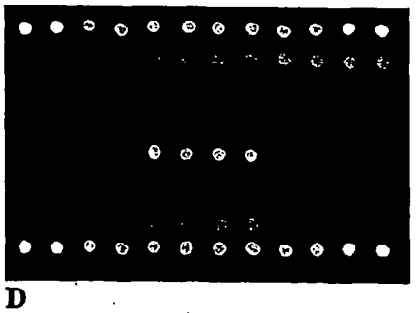
B


Calothrix
Spirulina
Woronichinia
Aphanizomenon
Microcystis
Universal
0
20000
40000
60000
80000

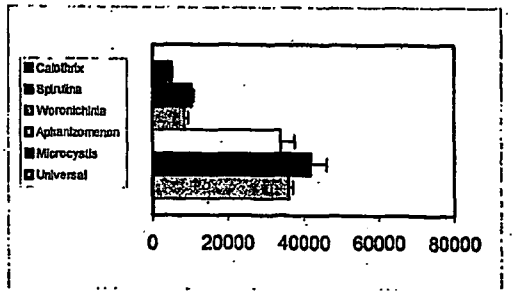


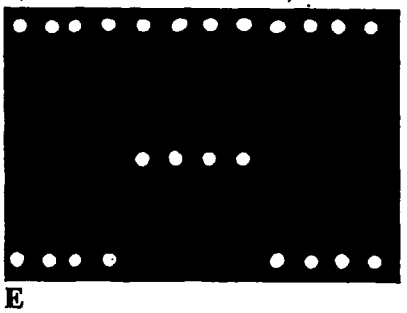
C


Calothrix
Spirulina
Woronichinia
Aphanizomenon
Microcystis
Universal
0
20000
40000
60000
80000



D


Calothrix
Spirulina
Woronichinia
Aphanizomenon
Microcystis
Universal
0
20000
40000
60000
80000



E

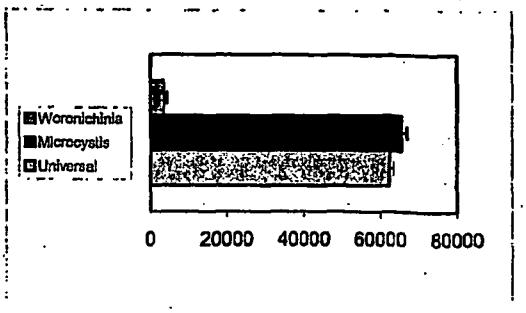
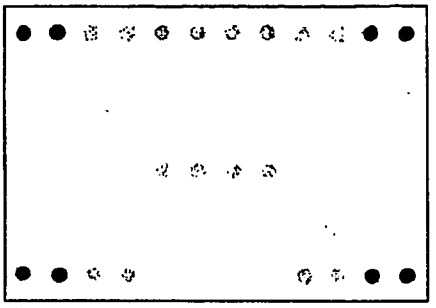
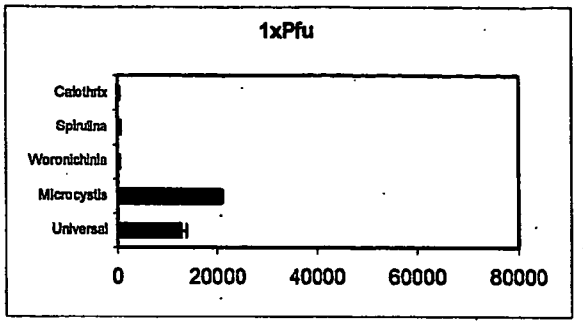
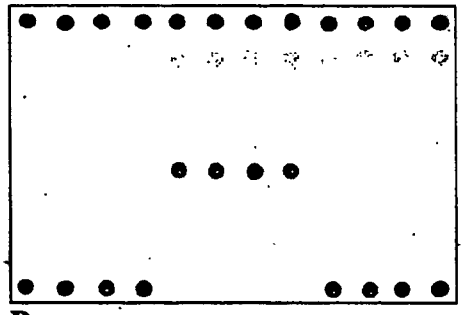

Woronichinia
Microcystis
Universal
0
20000
40000
60000
80000

Fig. 29



A


1xPfu
Calothrix
Spirulina
Woronichinia
Microcystis
Universal
0
20000
40000
60000
80000



B

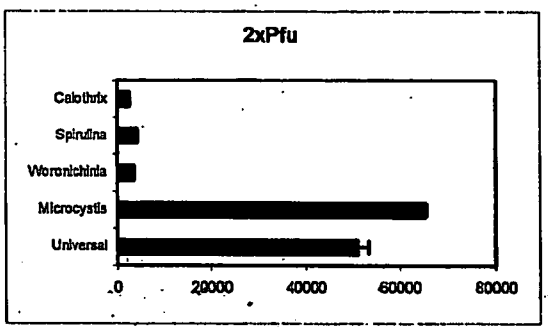

2xPfu
Calothrix
Spirulina
Woronichinia
Microcystis
Universal
0
20000
40000
60000
80000

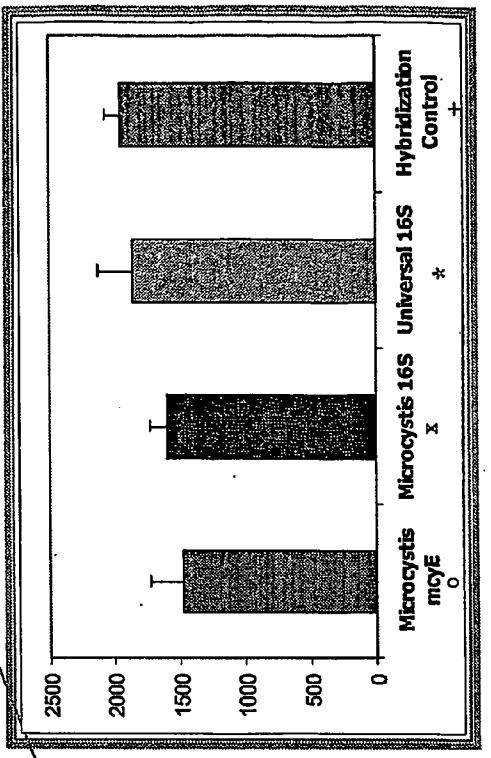
Fig. 30A

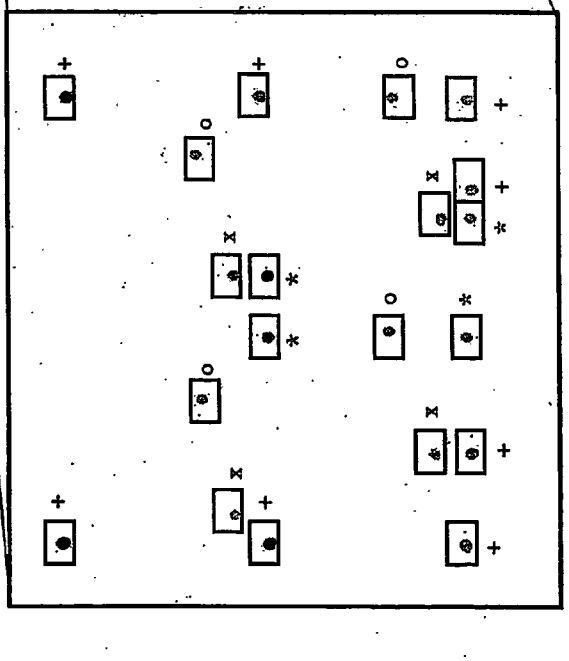
Fig. 30B

16S and *mcyE* discrimination onto Universal Array::an example

66	63	BLANK	1	2	3	4	5	1	2	3	4	5	BLANK	63	66
6	7	8	9	10	12	14	15	6	7	8	9	10	12	14	15
16	17	18	19	20	21	22	23	16	17	18	19	20	21	22	23
24	27	28	29	31	32	33	34	24	27	28	29	31	32	33	34
35	36	37	38	39	40	41	42	35	36	37	38	39	40	41	42
44	1B	3B	5B	11B	13B	15B	21B	44	1B	3B	5B	11B	13B	15B	21B
66	63	BLANK	1	2	3	23B	25B	23B	25B	1	2	3	BLANK	63	66
4	5	6	7	8	9	10	12	4	5	6	7	8	9	10	12
14	15	16	17	18	19	20	21	14	15	16	17	18	19	20	21
22	23	24	27	28	29	31	32	22	23	24	27	28	29	31	32
33	34	35	36	37	38	39	40	33	34	35	36	37	38	39	40
41	42	44	1B	3B	5B	11B	13B	41	42	44	1B	3B	5B	11B	13B
66	63	BLANK	66	15B	21B	23B	25B	15B	21B	23B	25B	66	BLANK	63	66

Zip Code 1B: Microcystis 16S rDNA
Zip Code 25: Universal 16S rDNA
Zip Code 40: Microcystis mcyE gene
Zip Code 66: Hybridization Control





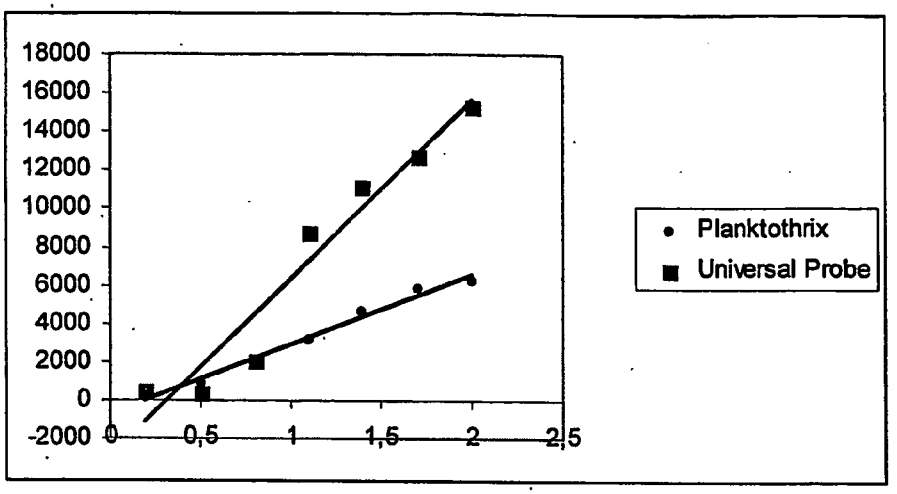
Microcystis 16S rDNA x

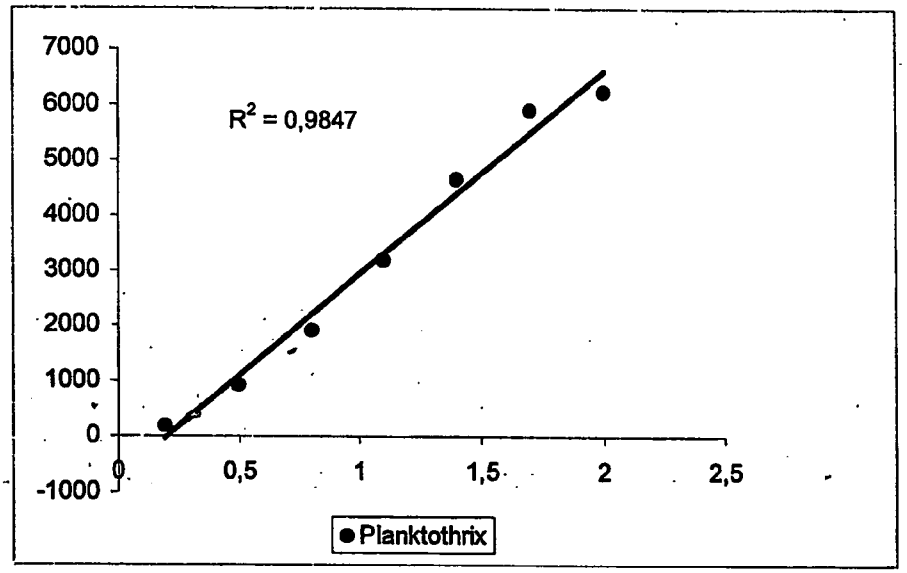
Universal 16S rDNA *

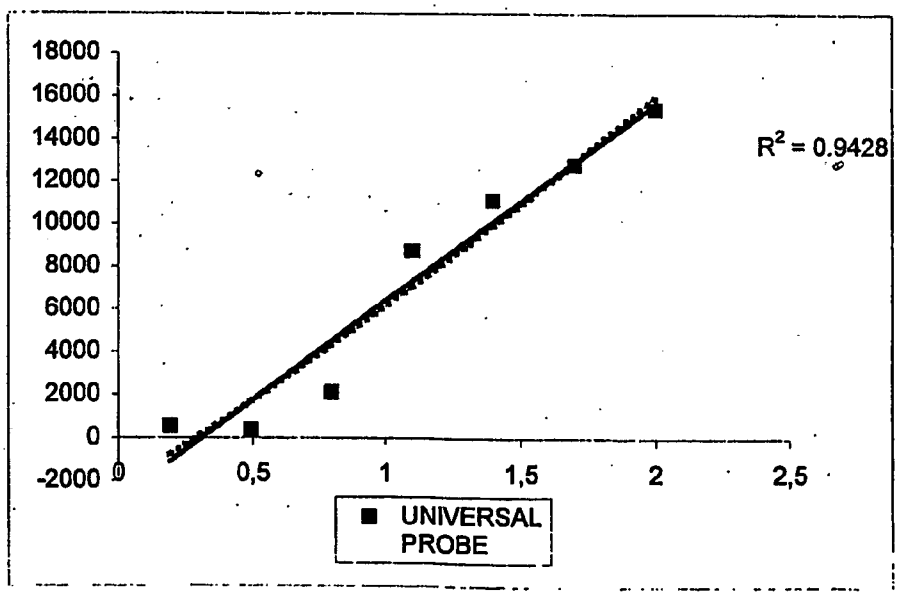
Microcystis mcyE gene o

Hybridization Control +

Fig. 31


18000
16000
14000
12000
10000
8000
6000
4000
2000
0
-2000
0
0,5
1
1,5
2
2,5
Planktothrix
Universal Probe


7000
6000
5000
4000
3000
2000
1000
0
-1000
0
0,5
1
1,5
2
2,5
R2 = 0,9847
Planktothrix


18000
16000
14000
12000
10000
8000
6000
4000
2000
0
-2000
0
0,5
1
1,5
2
2,5
R2 = 0.9428
UNIVERSAL PROBE

Polymorphism position	Group name	Zip code	SEQ ID NO: Discriminating probe	SEQ ID NO: Common probe	SEQ ID NO: C-zip code
657-A	MICROCYSTIS	1B	71 CGGTGGAAACTGGCAGACTAGAGA	91 GCAGTAGGGGTAGCAGGAATTCCC	111 GCTGAGGTCGATGCTGAGGTCGCA
841-A	PROCLOROCHOCCUS+ SYNECHOCOCCUS	3B	72 TGAACACTAGGTGTCGGGGGA	92 ATCGACCCCITCGGTGTCGTAG	112 GCTGCGATCGATGGTCAGGTGCTG
618-C	WORONICHINIA	5B	73 CAAGTCTGTCGTCAAAGAATGGAGC	93 TTAACTCCATAAAGGCIGTGGAAACTGAG	113 GCTGTACCCGATCGCAAGGTGGTC
1429-T	SPIRULINA	11B	74 ACACCATGGAAGCTGGCAACAT	94 CCGAAGTCGTTACTCCAACI	114 CGCAAGGTAGGTGCTGTACCCGCA
435-A	HALOTOLERANUS	13B	75 GGCTCTTGGGCTGTCAACCA	95 CTTTTCTCAGGGAAGAAGICCTGACGG	115 CGCACGATAGGTGGTCTACCGCTG
748-T	PROCLOROCHOCCUS MARINUS	15B	76 GAAGGCGCTCTGCTGGGCCATT	96 ACTGACGCTCATGGACGAAAGCC	116 CGCATACCAGGTCGCATACCGGTC
1051-C	OSCILLATORIA+ PLANKTOTHRIX	21B	77 CGTAAGGAACCAGAAGACAGGTGC	97 TGCATGGCTGTCGTCAGCTCGT	117 GGTCAGGTTACCGCTGCGATCGCA
747-C	NODULARIA	23B	78 GAAGGCGCTCTACTAGGCCGC	98 AACTGACACTGAGGGACGAAAGCTA	118 GGTCCGATTACCGGTCCGATGCTG
422-A	TRICHODESMIUM	27	79 GCGTGGGGGAGGAAIGCTCTA	99 GGGTTGTAAACCCCTTTTCTTTGGGAAG	119 GGGTATCCGTTCGGTGTTGCGTAGT
580-T	CYLINDROSPERMOPSIS	28	80 CGGAATGATTGGGCGTAAAGGGTCT	100 GCAGGTGGAACTGAAAGTCTGCTG	120 ACCTGGTCAATGGGACCATTGGTCC
632-G	CYLINDROSPERMUM	29	81 GTTAAAGAGCAAGGCTCAACCTTGTAAAG	101 GCAGTGGAAACTACATAGCTAGAGTGCG	121 TATGTCAGTGACGCGCTCAGCGTTG
1262-G	SYNECHOCYSTIS	31	82 GTCGGGACAACGGCAGCGAG	102 CTCGCGAGAGTAAGCGAATCCCA	122 TGGTGCTGGCGCAGACCTTTGTCTC
1328-C	NOSTOC	32	83 CCGGAGCTCAGTTCAGATCGC	103 AGGCTGCAACTCGCCTGC	123 ACCGCGCAAATGGACAGTGTGGCCA
484-G	PHORMIDIUM	33	84 AGAAAGTTGTGAAAGCAGCCTGACG	104 GTACCAGAGGAATCAGCATGGCTA	124 GACCCCAACTTGACACGTCGCAAGG
670-GA	LYNGBYA	34	85 GAACTAGAGGGCAGTAGGGGTAGA	105 GGGAATTCCCGGTGTAGCGGTG	125 GGAGAGTTTGGCGCGACCCTAACCT
801-C	GLOEOTHECE	35	86 TGTGCCGAAGCTAACGCGTTAAGTC	106 TCCCGCCTGGGGAGTACGCA	126 TGTGCTTACCGCACCTCGCAGTCGT
744-C	CALOTHRIX	36	87 GTGGCGAAAGCGTITITGCTAGGA	107 CAATACTGACACTGAGGGACGAAAGC	127 GTTGGGTATATCTCCCGGCGATCGC
857-C	LEPTOLYNGBYA	37	88 CGTATCGACCCGTGCAGTRCC	108 GTAGCTAACGCGTTAAGTITCCCGC	128 GTATTGGTGCTCGAGTCCGGCACGA
852-T	ANABAENA+APHANIZOMENON	38	89 GGIGTAGCTCGTATCGACCCGAGCT	109 GTRCCGIAGCTAACGCGTTAAGTATCCC	129 GTCTACGCCATCGCGGTGCTAAAGC
359-G	UNIVERSAL	25B	90 GACTCCTACGGGAGGCAGCAGTG	110 GGGAATTTTCCGCAATGGGC	130 GGTCTACCTACCCGCACGATGGTC

Fig. 32

Fig. 33A

Cyanobacterial strains used to validate the LDR procedure

Group	Strain	Geographic origin	Accession Number
ANABAENA/APHANIZOMENON	*Anabaena cylindrica* PCC 7122	Pond water, Cambridge, England	
	Anabaena sp. PCC 73105	Pond water, Cambridge, England	
	Anabaena sp. PCC 7108	Intertidal zone, California, U.S.A.	
	Anabaena sp. 90	Lake Vesijärvi, Finland	AJ133156
	Anabaena sp. 202A1	Lake Vesijärvi, Finland	AJ133159
	Aphanizomenon sp. 202	Lake Vesijärvi, Finland	AJ133153
	Aphanozomenon sp. PCC 7905	Lake Brielse Meer, The Netherlands	AJ133154
NOSTOC	*Nostoc* sp. PCC 7107	Shallow pond, California, U.S.A.	
	Nostoc sp. PCC 8114	Water bloom, Lake Hepetcon, New Jersey, U.S.A.	
	Nostoc punctiforme Hegewald 1971-108	Fish pond, Babat, Hungary	
	Nostoc linckia Hegewald 1971-144	Fish pond, Szeged, Feher-tó, Hungary	
	Nostoc sp. 152	Lake Sääksjärvi, Finland	AJ133161
MICROCYSTIS	*Microcystis aeruginosa* PCC 9354	Little Rideau Lake, Ontario, Canada	
	Microcystis sp. PCC 7005	Lake Mendota, Wisconsin, U.S.A.	
	Microcystis aeruginosa 1BB38S07	Bubano Basin, Imola, Italy	
	Microcystis aeruginosa 0BF29S03	Finissaggio Basin, Imola, Italy	
	Microcystis sp. 0BB35S01	Bubano Basin, Imola, Italy	
	Microcystis ichthyoblabe 0BB39S02	Bubano Basin, Imola, Italy	
	Microcystis wesenbergii NIES104	Freshwater lake, Japan	AJ133174
SYNECHOCOCCUS	*Synechococcus* sp. Hegewald 1974-30	Lake Kuusjärvi, Saukkolahti, Finland	
	Synechococcus sp. 0BB26S03	Bubano Basin, Imola, Italy	
	Synechococcus sp. WH 7803	Sargasso Sea	
	Synechococcus sp. WH 8103	Sargasso Sea	
	Synechococcus sp. 0BB42S04	Bubano Basin, Imola, Italy	
PROCHLOROCOCCUS MARINUS	*Prochlorococcus marinus* SS120	Sargasso Sea	
	Prochlorococcus marinus PCC 9511	Mediterranean Sea	
PLANKTOTHRIX	*Planktothrix* sp. 1LT27S08	Trasimeno Lake, Italy	
	Planktothrix sp. 2	Lake Markusbölefjärden, Finland	AJ133185
	Planktothrix sp. 28	Lake Markusbölefjärden, Finland	AJ133165
	Planktothrix sp. NIVA-CYA 126	Lake Långsjön; Finland	AJ133166
	Oscillatoria amphibia AGARDH Bai 1971-60	Pond Kakasszeg-tó, Hungary	
SPIRULINA	*Spirulina major* sp. PCC 6313	Brackish water, Berkeley, California, U.S.A.	
	Spirulina major 0BB22S09	Bubano Basin, Imola, Italy	

	Spirulina major 0BB36S18	Bubano Basin, Imola, Italy	
HALOTOLERANTS	*Cyanothece* sp. PCC 7418	Solar Lake, Israel	
NODULARIA	*Nodularia* sp. PCC73104/1	Alkaline soil, Spotted Lake, BC, Canada	AJ133184
	Nodularia sp. BY1	Baltic Sea	AJ133177
	Nodularia sp.NSPI-05	Coastal water, Peel Inlet, Australia	AJ133179
	Nodularia sp. HKVV	Baltic Sea	AJ133183
	Nodularia sp. NSOR-12	Coastal water, Oriental Lagoon, Tasmania, Australia	AJ133180
CYLINDROSPERMUM	*Cylindrospermum stagnale* PCC 7417	Soil, greenhouse, Sweden	AJ133163
SYNECHOCYSTIS	*Synechocystis* sp. PCC 6905	Low sality brine pond, Newark, California, U.S.A.	
	Synechocystis sp. PCC 7008	Shallow pond, Point Reyes Peninsula, California, U.S.A.	
CALOTHRIX	*Calothrix* sp. PCC 7714	Pool, India	AJ133164
	Calothrix marchica LEMM. Bai 1971-96	Pond Belsö-tó,Tihany, Hungary	
LEPTOLYNGBYA	*Leptolyngbya* sp. 0BB24S04	Bubano Basin, Imola, Italy	
	Leptolyngbya sp. 0BB30S02	Bubano Basin, Imola, Italy	
	Leptolyngbya sp. 0BB19S12	Bubano Basin, Imola, Italy	
	Leptolyngbya sp. 0BB32S02	Bubano Basin, Imola, Italy	
	Leptolyngbya SCHMIDLE Bai 1971-66	Fish pond, Szeged, Feher-tó, Hungary	
LYNGBYA	*Lyngbya* sp. 0BB32S04	Bubano Basin, Imola, Italy	
SNOWELLA	*Snowella litoralis* 0tu37s4	Lake Tuusulanjärvi, Finland	
	Snowella litoralis 0tu35s7		

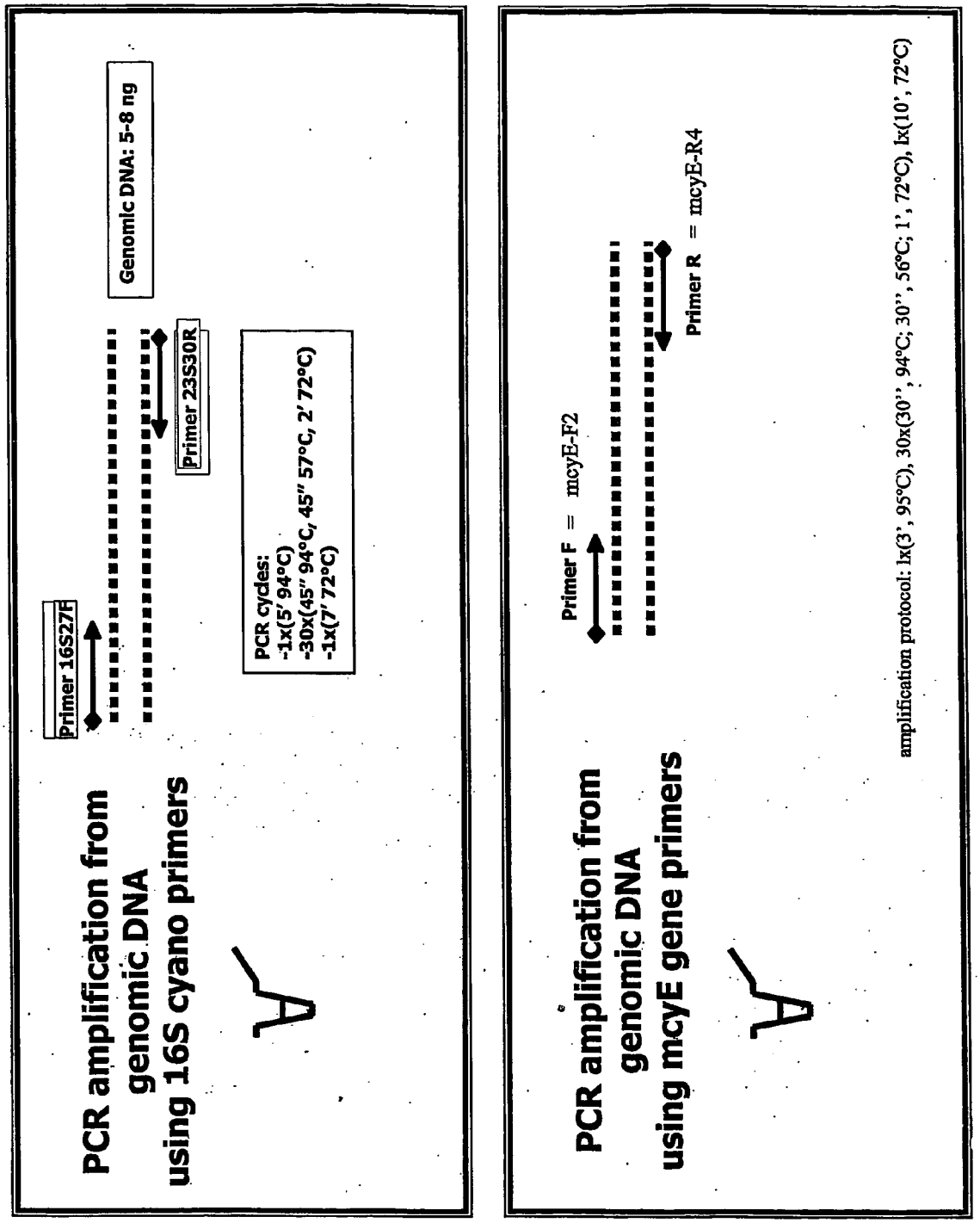
Fig. 33B

Table 3: Clones of 16S rRNA gene libraries obtained from environmental samples and used in the LDR reaction.

Group	Name	Environmental source
ANABAENA/APHANIZOMENON	OTU23C120	Lake Tuusulanjärvi (Finland)
	OTU23C167	Lake Tuusulanjärvi (Finland)
	OTU27CN57	Lake Tuusulanjärvi (Finland)
	OTU34C45	Lake Tuusulanjärvi (Finland)
	OTU34C47	Lake Tuusulanjärvi (Finland)
	OTU34C86	Lake Tuusulanjärvi (Finland)
	OTU34C109	Lake Tuusulanjärvi (Finland)
	OTU34C175	Lake Tuusulanjärvi (Finland)
	OES24F8	Lake Esch-sur-Sure (Luxembourg)
	OES24E16	Lake Esch-sur-Sure (Luxembourg)
MICROCYSTIS	OTU23C141	Lake Tuusulanjärvi (Finland)
	OTU27C97	Lake Tuusulanjärvi (Finland)
	OTU27CN214	Lake Tuusulanjärvi (Finland)
	OTU27CN235	Lake Tuusulanjärvi (Finland)
	OTU27CN255	Lake Tuusulanjärvi (Finland)
	OTU27CN258	Lake Tuusulanjärvi (Finland)
	OTU27CN297	Lake Tuusulanjärvi (Finland)
	OTU27CN318	Lake Tuusulanjärvi (Finland)
	OTU27CN324	Lake Tuusulanjärvi (Finland)
	OTU27CN329	Lake Tuusulanjärvi (Finland)
	OES46B58	Lake Esch-sur-Sure (Luxembourg)
SYNECHOCOCCUS	OTU34C70	Lake Tuusulanjärvi (Finland)
	OTU34C89	Lake Tuusulanjärvi (Finland)
	OTU34C113	Lake Tuusulanjärvi (Finland)
	OTU34C129	Lake Tuusulanjärvi (Finland)
	OTU34C134	Lake Tuusulanjärvi (Finland)
	OTU34C148	Lake Tuusulanjärvi (Finland)
	OTU34C154	Lake Tuusulanjärvi (Finland)
	OTU34C157	Lake Tuusulanjärvi (Finland)
	OTU34C176	Lake Tuusulanjärvi (Finland)
	OTU34C189	Lake Tuusulanjärvi (Finland)
PLANKTOTHRIX/OSCILLATORIA	OES28C14	Lake Esch-sur-Sure (Luxembourg))
	OES28C10	Lake Esch-sur-Sure (Luxembourg))
	OES28C20	Lake Esch-sur-Sure (Luxembourg))
	OES28D25	Lake Esch-sur-Sure (Luxembourg))
	OES28A2	Lake Esch-sur-Sure (Luxembourg))
	OES28C18	Lake Esch-sur-Sure (Luxembourg))
	OES28D3	Lake Esch-sur-Sure (Luxembourg))
	OES28C8	Lake Esch-sur-Sure (Luxembourg))
	OES28A5	Lake Esch-sur-Sure (Luxembourg))
WORONICHINIA	OES24A3	Lake Esch-sur-Sure (Luxembourg))
	OES46C21	Lake Esch-sur-Sure (Luxembourg))
	OES46C32	Lake Esch-sur-Sure (Luxembourg))
	OES46B48	Lake Esch-sur-Sure (Luxembourg))

Fig. 34

Fig. 35


PCR amplification from genomic DNA using 16S cyano primers
Primer 16S27F
Primer 23S30R
Genomic DNA: 5-8 ng
PCR cycles:
-1x(5' 94°C)
-30x(45" 94°C, 45" 57°C, 2' 72°C)
-1x(7' 72°C)
PCR amplification from genomic DNA using mcyE gene primers
Primer F = mcyE-F2
Primer R = mcyE-R4
amplification protocol: 1x(3', 95°C), 30x(30'', 94°C; 30'', 56°C; 1', 72°C), 1x(10', 72°C)

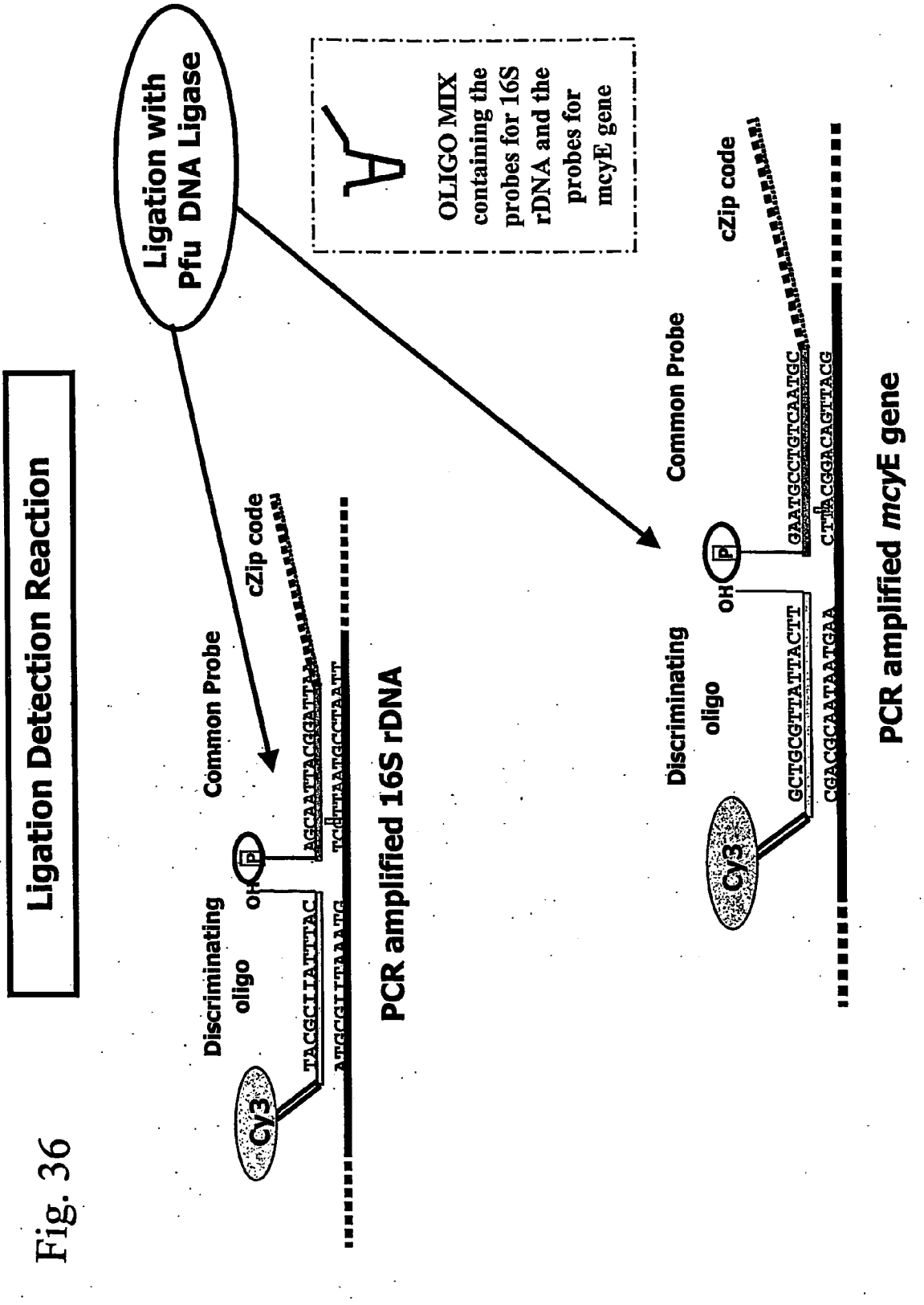

Ligation Detection Reaction
Ligation with Pfu DNA Ligase
OLIGO MIX containing the probes for 16S rDNA and the probes for mcyE gene
Discriminating oligo
Common Probe
cZip code
PCR amplified 16S rDNA
Discriminating oligo
Common Probe
cZip code
PCR amplified mcyE gene
Cy3
OH
P

Fig. 36

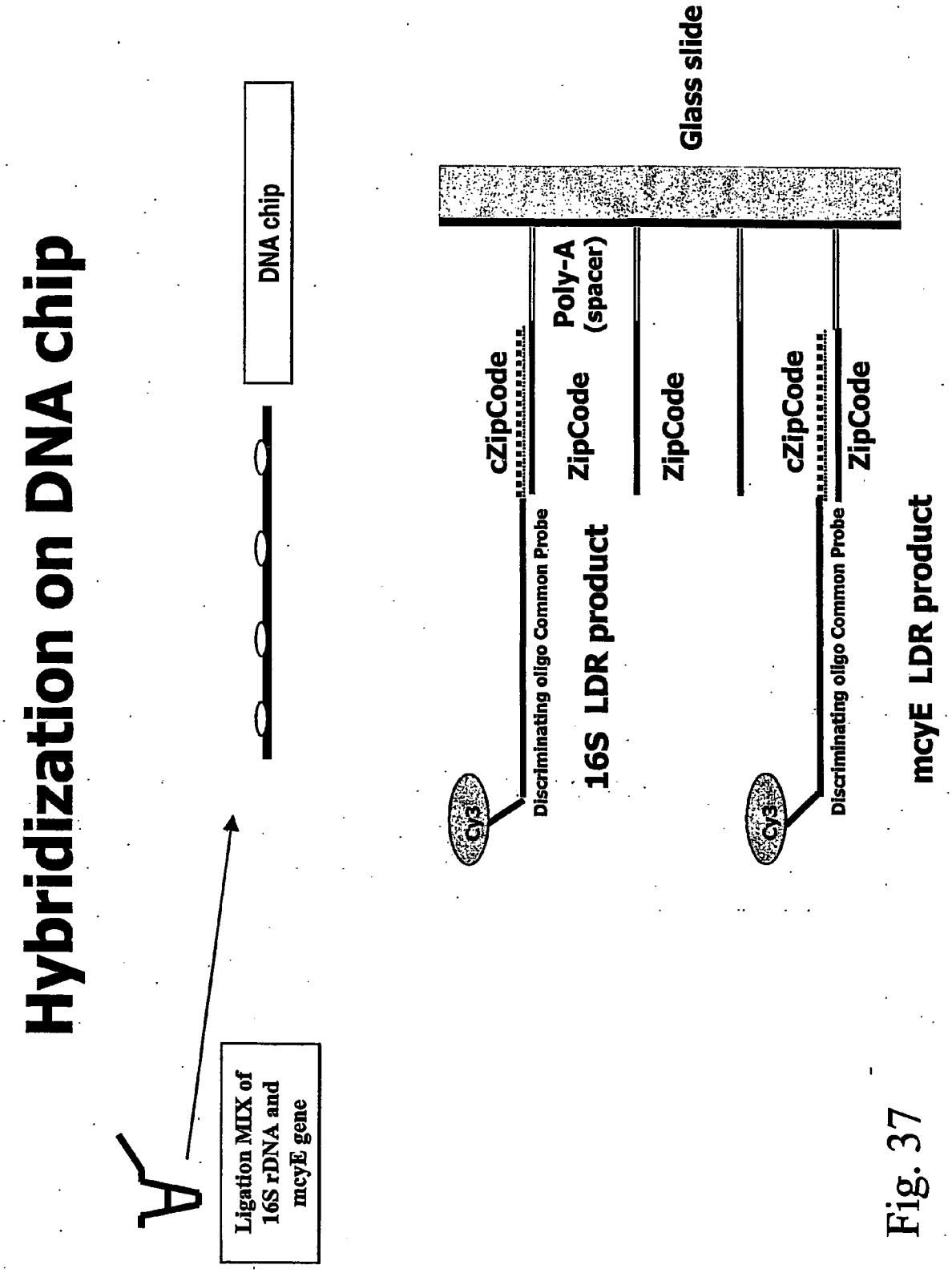

Hybridization on DNA chip
DNA chip
Ligation MIX of 16S rDNA and mcyE gene
Glass slide
cZipCode
Poly-A (spacer)
ZipCode
ZipCode
Discriminating oligo Common Probe
Cy3
16S LDR product
cZipCode
ZipCode
Discriminating oligo Common Probe
Cy3
mcyE LDR product

Fig. 37

Fig. 38A

>Anabaena 90 mcy D SEQ ID NO 131

CTCCAAAATCGACCTTTAATTATTGGTTCTGTCaAAACTAATTTAGGACA
TTTAGAAGGAGCAGCCGGAATTGCTGGGTTAATTAAAACTGTTTTAGCCC
TACAACATCACAAAATTCCTCCCCATCTTCACTTTCAACAACCCAACCCC
CGTTTTGATTGGAGTTCTCAGATTTTTGAAGTTCCAGTACATGGAAAAAA
CTGGCATCCTAGCCAACGAGAACGCATTGCTGGAGTAAGTTCTTTTGGAT
TTAGTGGTACTAATGCTCATATTATTGTTGGAGAAATTGCATCTAATTCT
CCACAGCCATCTGAGCAGAAATTTTACCTCCTGCCGCTTTCGGCTCGTTC
TCAAAAATCTCTCAAAGAATTAGCAAAAAATTATCAATACGCTTTAAATG
AATCTGTGAATTTCGCAGATACTTGTTTTACTGCCAGTACAGGAAGGGCT
ATTTTCCGGCATCGATTGTGTGTCTTGGCTGACTCAAATACTACAGCCGA
AAAAGCACTTGCTGATTTCCAAAAAGGTGAAGATTCTGATAATTTAATTA
CTCCAATTACATCAGAAACTCAAACAAAAGTAGTTTTCCTATTTTCAGGA
CAAGGTTCTCAATATTCAGGGATGGGACAAACTCTTTACAACCAAGAACC
CGTCTTTAAAAATACTCTGGAACTTTGTGACAACATTCTGCAACCTATTT
TAGGAAAGTCGCTCTTAGGTTTAATTTTTCAATTGCAAAATAGTGAACAG
CTAGAACAGACTCAAATCACACAACCAGCCCTGT

>Anabaena sp. 202A1/35 SEQ ID NO 132

CTCCAAAATCGACCTTTAATTATTGGTTCTGTCAAAACTAATTTAGGACA
TTTAGAAGGAGCAGCCGGAATTGCTGGGTTAATTAAAACTGTTTTAGCCC
TACAACATCACAAAATTCCTCCCCATCTTCACTTTCAACAACCCAACCCC
CGTTTTGATTGGAGTTCTCAGATTTTTGAAGTTCCAGTACATGGAAAAAA
CTGGCATCCTAGCCAACGAGAACGCATTGCTGGAGTAAGTTCTTTTGGAT
TTAGTGGTACTAATGCTCATATTATTGTTGGAGAAATTGCATCTAATTCT
CCACAGCCATCTGAGCAGAAATTTTACCTCCTGCCGCTTTCGGCTCGTTC
TCAAAAATCTCTCAAAGAATTAGCAAAAAATTATCAATACGCTTTAAATG
AATCTGTGAATTTCGCAGATACTTGTTTTACTGCCAGTACAGGAAGGGCT
ATTTTCCGGCATCGATTGTGTGTCTTGGCTGACTCAAATACTACAGCCGA
AAAAGCACTTGCTGATTTCCAAAAAGGTGAAGATTCTGATAATTTAATTA
CTCCAATTACATCAGAAACTCAAACAAAAGTAGTTTTCCTATTTTCAGGA
CAAGGTTCTCAATATTCAGGGATGGGACAAACTCTTTACAACCAAGAACC
CGTCTTTAAAAATACTCTGGAACTTTGTGACAACATTCTGCAACCTATTT
TAGGAAAGTCGCTCTTAGATTTAATTTTTCAATTGCAAAATAGTGAACAG
CTAGAACAGACTCAAATCACACAACCAGCCCTGT

>Anabaena sp. 66A SEQ ID NO 133

CCCCAAAATCAACCTTTAATTATTGGTTCTGTCAAAACTAATTTAGGACA
TTTAGAAGGAGCAGCCGGAATTGCTGGGTTAATTAAAACTGTTTTAGCCC
TACAACATCACAAAATTCCTCCCCATCTTCACTTTCAACAACCCAACTCC
CATTTTGATTGGAGTTCTCAGATTTTTGAAGTTCCAGTACATGGAAAAAA
CTGGCATCCTAGCCAACGAGAACGCATTGCTGGAGTAAGTTCTTTTGGAT
TTAGTGGTCCTAATGCTCATATTATTGTTGGAGAAATTGCATCTAATTCT
CCACAGCCATCTGAGCAGAAATTTTACCTCCTGCCGCTTTCGGCTCGTTC
TCAAAAATCTCTCAAAGAATTAGCAAAAAATTATCAATACGCTTTAAATG
AATCTGTGAATTTCGCAGATACTTGTTTTACTGCCAGTACAGGAAGGGCT
GTTTTCCAGCATCGATTGTGTGTCTTAGCTGACTCAAATACTACAGCCGA
AAAAGCACTTGCTGATTTCCAAAAAGGTGAAGATTCTGATAATTTAATTA
CTCCAATTACATCAGAAACTCAAACAAAAGTAGTTTTCCTATTTTCAGGA
CAAGGTTCTCAATATTCAGGGATGGGACAAACTCTTTACAACCAAGAACC
CGTCTTTAAAAATACTCTGGAACTTTGTGACAACATTCTGCAACCTATTT
TAGGAAAGTCGCTCTTAGATTTAATTTTTCAATTGCAAAATAGTGAACAG
CTAGAACAGACTCAAATCACACAACCAGCCCTGT

>Anabaena lemmermannii 202A2 SEQ ID NO 134

CTCCAAAATCGACCTTTAATTATTGGTTCTGTCAAAACTAATTTAGGACA
TTTAGAAGGAGCAGCCGGAATTGCTGGGTTAATTAAAACTGTTTTAGCCC
TACAACATCACAAAATTCCTCCCCATCTTCACTTTCAACAACCCAACCCC
CGTTTTGATTGGAGTTCTCAGATTTTTGAAGTTCCAGTACATGGAAAAAA
CTGGCATCCTAGCCAACGAGAACGCATTGCTGGAGTAAGTTCTTTTGGAT
TTAGTGGTACTAATGCTCATATTATTGTTGGAGAAATTGCATCTAATTCT
CCACAGCCATCTGAGCAGAAATTTTACCTCCTGCCGCTTTCGGCTCGTTC
TCAAAAATCTCTCAAAGAATTAGCAAAAAATTATCAATACGCTTTAAATG

Fig. 38B

AATCTGTGAATTTCGCAGATACTTGTTTTACTGCCAGTACAGGAAGGGCT
ATTTTCCGGCATCGATTGTGTGTCTTGGCTGACTCAAATACTACAGCCGA
AAAAGCACTTGCTGATTTCCAAAAAGGTGAAGATTCTGATAATTTAATTA
CTCCAATTACATCAGAAACTCAAACAAAAGTAGTTTTCCTATTTTCAGGA
CAAGGTTCTCAATATTCAGGGATGGGACAAACTCTTTACAACCAAGAACC
CGTCTTTAAAAATACTCTGGAACTTTGTGACAACATTCTGCAACCTATTT
TAGGAAAGTCGCTCTTAGATTTAATTTTTCAATTGCAAAATAGTGAACAG
CTAGAACAGACTCAAATCACACAACCAGCCCTGT

>Anabaena sp. 299 SEQ ID NO 135

CTCCAAAATCGACCTTTAATTATTGGTTCTGTCAAAACTAATTTAGGACA
TTTAGAAGGAGCAGCCGGAATTGCTGGGTTAATTAAAACTGTTTTAGCCC
TACAACATCACAAAATTCCTCCCCATCTTCACTTTCAACAACCCAACCCC
CGTTTTGATTGGAGTTCTCAGATTTTTGAAGTTCCAGTACATGGAAAAAA
CTGGCATCCTAGCCAACGAGAACGCATTGCTGGAGTAAGTTCTTTTGGAT
TTAGTGGTACTAATGCTCATATTATTGTTGGAGAAATTGCATCTAATTCT
CCACAGCCATCTGAGCAGAAATTTTACCTCCTGCCGCTTTCGGCTCGTTC
TCAAAAATCTCTCAAAGAATTAGCAAAAAATTATCAATACGCTTTAAATG
AATCTGTGAATTTCGCAGATACTTGTTTTACTGCCAGTACAGGAAGGGCT
ATTTTCCGGCATCGATTGTGTGTCTTGGCTGACTCAAATACTACAGCCGA
AAAAGCACTTGCTGATTTCCAAAAAGGTGAAGATTCTGATAATTTAATTA
CTCCAATTACATCAGAAACTCAAACAAAAGTAGTTTTCCTATTTTCAGGA
CAAGGTTCTCAATATTCAGGGATGGGACAAACTCTTTACAACCAAGAACC
CGTCTTTAAAAATACTCTGGAACTTTGTGACAACATTCTGCAACCTATTT
TAGGAAAGTCGCTCTTAGGTTTAATTTTTCAATTGCAAAATAGTGAACAG
CTAGAACAGACTCAAATCACACAACCAGCCCTGT

>Nostoc sp. 152 SEQ ID NO 136

CCCCAAAATCGCCCTTTAATTATTGGTTCTGTCAAAACTAATTTAGGGCA
TTTAGAAGGAGCAGCCGGAATTGCTGGGTTAATCAAAACTGTTTTAGCCT
TGCAACATCATAAAATTCCTCCGCATCTTCATTTTGAAAAACCCAATCCC
CGTTTTGATGGGAGTTCTCATATTTTTGAAGTTCCAGTACATGGTAAAAA
CTGGCATCCTAGCGAACGAGAAAGAATTGCGGGTGTGAGTTCCTTCGGAT
TTAGTGGTACGAATGCCCATGTTATTGTGGGAGAAATTGCATCTAATTTT
TCACAACAATCTGAGCATCAGCTTTACCTTTTGCCTCTTTCGGCTCGTTC
TGAAAAGTCCCTCAAAGAGTTAGCCAAAAATTATCAATCTGCTTTAAATG
AATCTGTAAATTTAGCCGATGCTTGTTTTACCGCTAGTACAGGAAGGGCT
AATTTTCGGCATCGATTGTGTATTCTAGCTGACTCAATCACCACAGCAGA
GAAAGCGCTTACTGATTTCCAGAAAGGTGAGGATTCTGAGCATATAATTA
CGCAAATTGCATCAGAAACTCAACCAAATATAGCTTTACTATTCTCAGGA
CAAGGTTCTCAATATTCCGGGATGGGACAAACTCTTTACAACAAAGAACC
TGTCTTTAAAAATACTCTAGATATTTGTGACCAAATCTTGCAACCTATTT
TAGGAGCATCGCTATTAGATTTAATCTTTGAAGTGTCGAATAGCGATTTG
CTCGAACAAACTCAAATCACACAGCCAGCGCTCT

>Nodularia sp. F81 SEQ ID NO 137

CCCCAAAATCGACCTTTAATAATNGGTTCCGTCAAAACCAATTTAGGACA
TTTAGAAGGAGCAGCCGGAATAGCTGGGTTAATCAAAACTGTTTTAGCCT
TGCAAAAGCATCAAATTCCTCCCCATCTCCACTTTCAGCAACCAAACCCC
CGTTTTGATTGGAGTACCGATATTTTTGAAGTCCCAGTCCATGGAAAAAA
TGGGTATCCTAGCCAGCGAGAACGCATTGCCGGCGTGAGTTCTTTGGGAT
TTAGGGGTACTAATGCCCATATTATGGTCGGAGAAGTTGCCCGAAATTCT
CCCCAACCATCAGAGCCAAAATTTTACTTACTACCCTTTTCCGCTCGTTC
CGTAACATCTCTGCAACAGTTAGCTAAAAATTATCAGTCCGCCTTGAATG
ACTCTGTTAATTAAGCAGATGCTGGTTTTACTATCAGCACAGGTAGGGCT
ATTTTCCCCCATCGATTATGTATTCTCGCCGACTCCATTTCCACAGCCCA
AAAAGCCCTGGCTGACTTCCAAACAGGTGAAGATTCAGACAATTTAATTA
CACAAATTAGCTCAGAAACTCAGCCCAAGATCGCCTTTCTCTTCACTGGA
CAGGGTTCTCAATATACCGGCATGGGAGAAACGCTTTACAACCAAGAACC
AGTCTTTAGAAATACTCTAGATATTTGTGACCAAATCCTCCAACCAATTT
TAGGAACATCACTATTAGATTTAATCTGGCAATTGCCAAATAGCGAATTA
CTGGCACAAACGCAAATCACACAGCCAGCCCTAT

>Nodularia spumigena HEM SEQ ID NO 138

Fig. 38C

CCCCAAAATCGACCTTTAATAATNGGTTCCGTCAAAACCAATTTAGGACA
TTTAGAAGGAGCAGCCGGAATAGCTGGGTTAATCAAAACTGTTTTAGCCT
TGCAAAAGCATCAAATTCCTCCCCATCTCCACTTTCAGCAACCAAACCCC
CGTTTGGATGGGAGTACCGATATTTTGGAAGTCCCAGTACATGGAAAAAA
TGGGTATCCTAGCCAGCGAGAACGCATGGCCGGCGTGAGTTCTTTGGGAT
TTAGGGGTACTAATGCCCATATTATGGTCGGAGAAGTTGCCCGAAATTCT
CCCCAACCATCAGAGCCAAAATTTTACTTACTACCCTTTTCCGCTCGTTC
CGTAACATCTTTGCAACAGTTAGCTAAAAATTATCAGTCCGCCTTGAATG
ACTCTGTTAATTAAGCAGATGCTGGTTTTACTATCAGCACAGGTAGGGCT
ATTTTCCCCCATCGATTATGTATTCTCGCCGACTCCATTTCCACAGCCAA
AAAAGCCCTGGCTGACTTCCAAACAGGTGAAGATTCAGACAATTTAATTA
CACAAATTAGCTCAGAACCTCAGCCCAAGATCGCCTTTCTCTTCACTGGA
CAAGGTTCTCAATATACCGGCATGGGAGAAACGCTTTACAACCAAGAACC
AGTCTTTAGAAATACTCTAGATATTTGTGACCAAATCCTCCAACCAATTT
TAGGAACATCACTATTAGATTTAATCTGGCAATTGCCAAATAGCGAATTA
CTGGCACAAACGCAAATCACACAGCCAGCCCTAT

>Nodularia spumigena BY1 SEQ ID NO 139

CCCCAAAATCGACCTTTAATAATNGGTTCCGTCAAAACCAATTTAGGACA
TTTAGAAGGAGCAGCCGGAATAGCTGGGTTAATCAAAACTGTTTTAGCCT
TGCAAAAGCATAAAATTCCTCCCCATCTCCACTTTCAGCAACCAAACCCC
CGTTTTGATGGGAGCCCCGATATTTTGGAAGTCCCAGTCCATGGAAAAAA
TGGGTATCCTAGCCAGCGAGAACGCATGGCCGGCGTGAGTTCTTTGGGAT
TTAGGGGTACTAATGCCCATATTATGGTCGGAGAAGTTGCCCGAATTTCT
CCCCAACCATCAGAGCCAAAATTTTACTTACTACCCTTTTCCGTTCGTTC
CGTAACATCTTTGCAACAGTTAGCTAAAAATTATCAGTCCGCCTTGAATG
ACTCTGTTAATTAAGCAGATGCTGGTTTTACTATCAGCACAGGTAGGGCT
ATTTTCCCCCATCGATTATGTATTCTCGCCGACTCCATTTCCACAGCCAA
AAAAGCCCTGGCTGACTTCCAAACAGGTGAAGATTCAGACAATTTAATTA
CACAAATTAGCTCAGAAACTCAGCCCAAGATCGCCTTTCTCTTCACTGGA
CAAGGTTCTCAATATACCGGCATGGGAGAAACGCTTTACAACCAAGAACC
AGTCTTTAGAAATACTCTAGATATTTGTGACCAAATCCTCCAACCAATTT
TAGGAACATCACTATTAGATTTAATCTGGCAATTGCCAAATAGCGAATTA
CTGGCACAAACGCAAATCACACAGCCAGCCCTAT

>Microcystis aeruginosa PCC7941 SEQ ID NO 140

TCCCCTAATCGGCCCTTAATTATTGGTTCAGTAAAAACGAATTTAGGCCA
TTTAGAAGGTGCTGCTGGAATTGCCGGATTAATTAAAACCGTTCTAGCGT
TACAACACCATAAAATACCTCCTCATCTTCACTTTAAAAATCCTAATCCC
CGCTTTGATTGGAGTTCTCATATTTTTGAAGTTCCTGTACAAGGAAAACC
TTGGAATATCAGCGAACGTCCAAGGATTGCTGGAGTAAGTTCCCTTGGAT
TTAGTGGAACGAATGCTCATATCATTGTTGGGGAAATTGATGCTGATTTA
CCTCAAGCTTCGGAGAATAATTTTTATCTATTACCCCTTTCGGCTCGTTC
TGAACAATCCCTGCAAGAGTTAGCGAGAAGCTATCAAGATATTTTGACTG
AGTCTATCAATTTAGCTGATGTTTGTTTTACCACCAGTACAGGGCGGGGG
ATTTTTCCGCAACGAATCTGTATTTTAGCAGACTCAATAACGACGGCACA
ACGAGCATTAATTGATTACCAAGATGGTGAAGATTCTGACTCATTAATCC
GACCGATTTTATCAGAGACTCCGCCAAAGATGGCTTTCCTCTTTTCTGGA
CAAGGCTCTCAATATTCTGGCATGGGAGAAACTCTTTATAACCGAGAAGT
TGTTTTTAAGGAAACATTAGATCTTTGTGATCAAATTTTAGAACCCCTTT
TAGAAAAATCTCTCCTAGATTTAATCTTTCAAGAGCAAAATAGCCAGTTA
TTAGAGGAAACCCAAATCACTCAGCCGGTGATTT

>Microcystis aeruginosa NIES 89 SEQ ID NO 141

TCCCCTAATCGGCCTTTAATTATTGGTTCAGTAAAAACGAATTTAGGCCA
TTTAGAAGGTGCTGCTGGAATTGCCGGATTAATTAAAACCGTTCTAGCGT
TACAACACCATAAAATACCTCCTCATCTTCACTTTAAAAATCCTAATCCC
CGCTTTGATTGGAGTTCTCATATTTTTGAAGTTCCTGTACAAGGAAAACC
TTGGTATATCAGCGAACGTTCAAGGATTGCTGGAGTAAGTTCCTTTGGAT
TTAGTGGAACGAATGCTCATATCATTGTTGGGGAAATTGACGCTGATTTA
CCTCAACCTTCGGAGAATAATTTTTATCTATTACCCCTTTCGGCTCGTTC
TGAAAAATCGCTGCAAGAGTTAGCGAGAAGTTATCAAGATATTTTGACTG
AGTCTATCAATTTAGCTGATGTTTGTTTTACGGCCAGTACAGGGCGGGGG

Fig. 38D

ATTTTTCCGCAACGAATCTGTATTTTAGCAGACTCAATAGCCACGGCACA
ACGAGCATTAATTGATTACCAAGATGGTGAAGATTCTGACTCATTAATCC
GACCGATTTTATCAGAGACTCCGCCAAAGATAGCTTTCCTCTTTTCTGGA
CAAGGCTCTCAATATTCCGGCATGGGAGAAACTCTTTATAACCGAGAAGT
TGTTTTTAAGGAAACATTAGATCTTTGTGATCAAATTTTAGAACCCCTTT
TAGAAAA-TCTCTCCTAGATTTAATCTTTCAAGAGCAAAATAGCCAGTTA
TTAGAGGAAACCCAAATCACTCAGCCGGTGATTT

>Microcystis viridis NIES102 SEQ ID NO 142

TCCCCTAATCGGCCTTTAATTATTGGTTCAGTAAAAACGAATTTAGGCCA
TTTAGAAGGTGCTGCTGGAATTGCCGGATTAATTAAAACCGTTCTAGCGT
TACAACACCATAAAATACCTCCTCATCTTCACTTTAAAAATCCTAATCCC
CGCTTTGATTGGAGTTCTCATATTTTTGAAGTTCCTGTACAAGGAAAACC
TTGGGATATCAGCGAACGTTCAAGGATTGCTGGAGTAAGTTCCTTTGGAT
TTAGTGGAACGAATGCTCATATCATTGTTGGGGAAATTGACGCTGATTTA
CCTCAACCTTCGGAGAATAATTTTTATCTATTACCCCTTTCGGCTCGTTC
TGAAAAATCGCTGCAAGAGTTAGCGAGAAGTTATCAAGATATTTTGACTG
AGTCTATCAATTTAGCTGATGTTTGTTTTACCGCCAGTACAGGGCGGGGG
ATTTTTCCGCAACGAATCTGTATTTTAGCAGACTCAATAGCCACGGCACA
ACGAGCATTAATTGATTACCAAGATGGTGAAGATTCTGACTCATTAATCC
GACCGATTTTATCAGAGACTCCGCCAAAGATAGCTTTCCTCTTTTCTGGA
CAAGGCTCTCAATATTCCGGCATGGGAGAAACCCTTTATAACCGAGAAGT
TGTTTTCAAGGAAACATTAGATCTTTGTGATCAAATTCTAGAACCCCTTT
TAGAAAAATCTCTCTTAGATTTAATCTTTCAAGAGCAAAATAGTGAGTTA
TTAGAGGAAACCCAAATCACTCAGCCGGTGATTT

>Microcystis sp. K139 SEQ ID NO 143

TCCCCTAATCGGCCCTTAATTATTGGTTCAGTAAAAACGAATTTAGGCCA
TTTAGAAGGTGCTGCCGGAATTGCCGGATTAATTAAAACCGTTCTAGCGT
TACAACACCATAAAATACCTCCTCATCTTCACTTTAAAAATCCTAATCCC
CGCTTTGATTGGAGTTCTCATATTTTTGAAGTTCCTGTACAAGGAAAACC
TTGGGATATCAGCGAACGTCCAAGGATTGCTGGAGTAAGTTCCTTTGGAT
TTAGTGGAACGAATGCTCATATCATTGTTGGGGAAATTGATGCTGATTTG
CCTCAAGCTTCGGAGAATAATTTTTATCTATTACCCCTTTCGGCTCGTTC
TGAACAATCCCTGCAAGAGTTAGCGAGAAGTTATCAAGATATTTTGACTG
AGTCTATCAATTTAGCTGATGTTTGTTTTACCACCAGTACAGGGCGGGGG
ATTTTTCCGCAACGAATCTGTATTTTAGCAGACTCAATAACCACGGCACA
ACGAGCATTAATTGATTACCAAGATGGTGAAAATTCTGACTCATTAATCC
GACCGATTTTATCAGAGACTCCACAAAAGATAGCTTTCCTCTTTTCTGGA
CAAGGCTCTCAATATTCCGGCATGGGAGAAACCCTTTATAACCGAGAAGT
TGTTTTCAAGGAAACATTAGATCTTTGTGATCAAATTTTAGAACCCCTTT
TAGAAAAATCTCTCTTAGATTTAATCTTTCAAGAGCAAAATAGTGAGTTA
TTAGAGGAAACCCAAATCACTCAGCCGGTGATTT

>Microcystis aeruginosa PCC7806 SEQ ID NO 144

TCCCCTAATCGGCCCTTAATTATTGGTTCAGTAAAAACGAATTTAGGCCA
TTTAGAAGGTGCTGCCGGAATTGCCGGATTAATTAAAACCGTTCTAGCGT
TACAACACCATAAAATACCTCCTCATCTTCACTTTAAAAATCCTAATCCC
CGCTTTGATTGGAGTTCTCATATTTTTGAAGTTCCTGTACAAGGAAAACC
TTGGAATATCAGCGAACGTCCAAGGATTGCTGGAGTAAGTTCCTTTGGAT
TTAGTGGAACGAATGCTCATATCATTGTTGGGGAAATTGATGCTGATTTG
CCTCAAGCTTCGGAGAATAATTTTTATCTATTACCCCTTTCGGCTCGTTC
TGAACAATCCCTGCAAGAGTTAGCCAGAAGCTATCAAGATATTTTGACTG
AGTCTATCAATTTAGCTGATGTTTGTTTTACCACCAGTACAGGGCGGGGG
ATTTTTCCGCAACGAATCTGTATTTTAGCAGACTCAATAACGACGGCACA
ACGAGCATTAATTGATTACCAAGATGGTGAAGATTCTGACTCATTAATCC
GACCGATTTTATCAGAGACTCCGCCAAAGATATCTTTCCTCTTTTCTGGA
CAAGGCTCTCAATATTCTGGCATGGGAGAAACTCTTTATAACCGAGAAGT
TGTTTTTAAGGAAACATTAGATCTTTGTGATCAAATTTTAGAACCCCTTT
TAGAAAAATCTCTCCTAGATTTAATCTTTCAAGAGCAAAATAGCCAGTTA
TTAGAGGAAACCCAAATCACTCAGCCGGTGATTT

>Microcystis sp. 205 SEQ ID NO 145

TCCCCTAATCGGCCTTTAATTATTGGTTCAGTAAAAACGAATTTAGGCCA

Fig. 38E

TTTAGAAGGTGCTGCCGGAATTGCTGGATTAATTAAAACAGTTCTAGCTT
TACAACACCATAAAATACCTCCTCATCTTCACTTTAAAAATCCTAATCCC
CGCTTTGATTGGAGTTCTCATATTTTTGAAGTTCCTGTACAAGGAAAACC
TTGGAATATCAGCGAACGTGCAAGGATTGCTGGAGTAAGTTCCTTTGGAT
TTAGTGGAACGAATGCTCATATCATTGTTGGGGAAATTGATGCTGATTTG
CCTCAAGCTTCGGAGAATAATTTTTATCTATTACCCCTTTCGGCTCGTTC
TGAACAATCCCTGCAAGAGTTAGCGAGAAGCTATCAAGATATTTTGACTG
AGTCTATCAATTTAGCTGATGTTTGTTTTACCACCAGTACAGGGCGGGGG
ATTTTTCCGCAACGAATCTGTATTTTAGCAGACTCAATAACGACGGCACA
ACGAGCATTAATTGATTACCAAGATGGTGAAGATTCTGACTCATTAATCC
GACCGATTTTATCAGAGACTCCGCCAAAGATAGCTTTCCTCTTTTCTGGA
CAAGGCTCTCAATATTCTGGCATGGGAGAAACTCTTTATAACCGAGAAGT
TGTTTTTAAGGAAACATTAGATCTTTGTGATCAAATTTTAGAACCCCTTT
TAGAAAAATCTCTCCTAGATTTAATCTTTCAAGAGCAAAATAGCCAGTTA
TTAGAGGAAACCCAAATCACTCAGCCGGTGATTT

>Planktothrix sp. NIVA-CYA 127 SEQ ID NO 146

CCTCCGAATCAACCTTTAGTTATTGGTTCTGTCAAAACAAATTTAGGACA
CTTAGAAGGAGCAGCAGGAATTGCCGGATTAATTAAAACAGTTTTAGCTT
TACAACATCATAAAATTCCTCCCCATCTTCACTTTAAAAAACCCAACCCT
CGGTTGGATGGGAGTTCTAATATTTTTGAAGTTCCTGTAGGCGGAAAACC
CTGGAATCCCAGTGAACGCCAAAGAATTGCCGGGGTAAGTTCCTTTGGCT
TTAGTGGAACAAATGCTCATATTATTGTGGGAGAAATTGACTCTAGTTTA
CCTAAAAAATCTGAGCCTAACTTTTACCTATTACCGCTTTCGGCTCGTTC
TGAAAAATCTCTCCAAGAGTTAACTAAAAATTATCAAAATGCTTTGAATG
GGTCTGTCAATTTTGCTGATGTTGGTTTTACGGCTACTACAGGACGGGCT
ATTTTTCAGCATCGAATATGTATTTTAGCTGAATCAATGACAACAGCACA
GGCAGCACTGGTTAGTTTCCAAAAAGGTGAAAATTCTCAACATTTAATTA
CACCAATTTTATCAGAAAACAAGCTAAAAATAGCTTTTCTATTTTCAGGA
CAAGGCTCACAATATTCGGAAATGGGAGAAACCCTTTATCACCGAGAACC
TGTCTTTAAAAATACTTTAGATATTTGTAATGAAATCCTAGAACCTATTT
TAGAAAAATCCCTGTTAGATTTAATCTTTAAATTGCCCAATAGCCAGCTA
TTAGAACAGACTCAAATCACCCAGCCCGTGCTAT

>Planktothrix sp. NIVA-CYA 128/R SEQ ID NO 147

CCTCCGAATCAACCTTTAGTTATTGGTTCTGTCAAAACAAATTTAGGACA
CTTAGAAGGAGCAGCAGGAATTGCCGGATTAATTAAAACAGTTTTAGCTT
TACAACACCATAAAATTCCTCCCCATCTTCACTTTAAAAAACCCAACCCT
CGGTTGGATGGGAGTTCTAATATTTTTGAAGTTCCTGTAGGCGGAAAAAC
CTGGAATCCCAGTGAACGCCAAAGAATTGCCGGGGTAAGTTCCTTTGGCT
TTAGTGGAACAAATGCTCATATTATTGTGGGAGAAATTGACTCTAGTTTA
CCTAAAAAATCTGAGCCTAACTTTTACCTATTACCGCTTTCGGCTCGTTC
TGAAAAATCTCTCCAAGAGTTAACTAAAAATTATCAAAATGCTTTGAATG
GGTCTGTCAATTTTGCTGATGTTGGTTTTACGGCTACTACAGGACGGGCT
ATTTTTCAGCATCGAATATGTATTTTAGCTGAATCAATGACAACAGCACA
GGCAGCACTGGTTAGTTTCCAAAAAGGTGAGGATTCTCAACATTTAATTA
CACCAATTTTATCAGAAAACAAGCTAAAAATAGCTTTTCTATTTTCAGGA
CAAGGATCGCAATATTCAGGAATGGGAGAAACCCTTTATCACCGAGAACC
TGTCTTTAAAAATACTTTAGATCTTTGTAATGAAATCCTAGAACCTATTT
TAGAAAAATCCCTGTTAGATTTAATCTTTAAATTGCCCAATAGCCAGCTA
TTAGAACAGACTCAAATCACCCAGCCCGTGCTAT

>Planktothrix 49 SEQ ID NO 148

CCTCCGAATCAACCTTTAGTTATTGGTTCTGTCAAAACAAATTTAGGACA
CTTAGAAGGAGCAGCAGGAATTGCCGGATTAATTAAAACAGTTTTAGCTT
TACAACATCATAAAATTCCTCCCCATCTTCACTTTAAAAAACCCAACCCT
CGGTTTGATGGGAGTTCTAATATTTTTGAAGTTCCTGTAGGCGGAAAACC
CTGGAATCCCAGTGAACGCCAAAGAATTGCCGGGGTAAGTTCCTTTGGCT
TTAGTGGAACAAATGCTCATATTATTGTGGGAGAAATTGACTCTAGTTTA
CCTAAAAAATCTGAGCCTAACTTTTACCTATTACCGCTTTCGGCTCGTTC
TGAAAAATCTCTCCAAGAGTTAACTAAAAATTATCAAAATGCTTTGAATG
GGTCTGTCAATTTTGCTGATGTTTGTTTTACGGCTACTACAGGACGGGCT
ATTTTTCAGCATCGAATATGTATTTTAGCTGAATCAATGACAACAGCCCA

Fig. 38F

AGCAGCACTGGTTAGTTTCCAAAAAGGTGAAAATTCTCAACAATTAATTA
CACCAATTTTATCAGAAAACAAGCTAAAAATAGCTTTTCTATTTTCAGGA
CAAGGCTCACAATATTCGGGAATGGGAGAAACCCTTTATCACCGAGAACC
TGTCTTTAAAAATACTTTAGATATTTGTAATGAAATCCTAGAACCTATTT
TAGAAAAATCCCTGTTAGATTTAATCTTTAAATTGCCCAATAGCCAGCTA
TTAGAACAGACTCAAATCACCCAGCCCGTGCTAT

>Planktothrix sp. NIVA-CYA 126/8 SEQ ID NO 149

CCTCCGAATCAACCTTTAGTTATTGGTTCTGTCAAAACAAATTTAGGACA
CTTAGAAGGAGCAGCAGGAATTGCCGGATTAATTAAAACAGTTTTAGCTT
TACAACACCATAAAATTCCTCCCCATCTTCACTTTAAAAAACCCAACCCT
CGGTTTGATTGGAGTTCTAATATTTTTGAAGTTCCTGTAGGCGGAAAAAC
CTGGAATCCCAGTGAACGCCAAGGAATTGCCGGGGTAAGTTCCTTTGGCT
TTAGTGGAACAAATGCTCATATTATTGTGGGAGAAATTGACTCTAGTTTA
CCTAAAAAATCTGAGCCTAACTTTTACCTATTACCGCTTTCGGCTCGTTC
TGAAAAATCTCTCCAAGAGTTAACTAAAAATTATCAAAATGCTTTGAATG
GGTCTGGCAATTTTGCTGATGTTTGTTTTACGGGTACTACAGGACGGGCT
ATTTTTCAGCATCGAATATGTATTTTAGCTGAATCAATGACAACAGCACA
AGCAGCACTGGTTAGTTTCCAAAAAGGTGAGGATTCTCAACAATTAATTA
CACCAATTTTATCAGAAAACAAGCTAAAAATAGCTTTTCTATTTTCAGGA
CAAGGATCGCAATATTCAGGAATGGGAGAAGCCCTTTATCACCGAGAACC
TGTCTTTAAAAATACTTTAGATCTTTGTAATGAAATCCTAGAACCTATTT
TAGAAAAATCCCTGTTAGATTTAATCTTTAAATTGCCCAATAGCCAGCTA
TTAGAACAGACTCAAATCACCCAGCCCGTGCTAT

Group name	Zip Code	Discriminating probe sequence (5'-3')	SEQ ID NO	Common probe sequence (5'-3')	SEQ ID NO
SNOWELLA	*10*	GCGTAAAGAGTCCGTAGGTGGAAG	150	TTC AAG TCT GCG GTC AAA GAA TGG AGG	157
TOXIC ANABAENA	*12*	GATGTGATGCGAATCTCATAAACCGG	151	AGC TCA GTT CAG ATC GAA GGC TGC G	158
APHANIZOMENON CLUSTER I	*14*	GCGAAAGCGTTCTGCTAGACCTGT	152	ACT GAC ACT GAG GGA CGA AAG CTA GG	159
APHANIZOMENON CLUSTER II	*15*	CCAAGGCTTGACATGTCACGAATCCTA	153	TTG AAA GAT GGG AGT GCC TTC GGG AT	160
ANABAENA CLUSTER I	*16*	GAGCAGTGGAAACTACAAAGCTAGAGTT	154	TGG TCG GGG CAG AAG GAA TTC CTC	161
ANABAENA CLUSTER II	*17*	TCACGCTCAACGTGATCAAGGCG	155	GTG GAA ACT ACA AAG CTA GAG TGT G	162
ANABAENA+APHANIZOMENON SUBCLUSTER	*18*	CATGTCACGAATTCCGTTGAAAGATGGA	156	AGT GCC TTC GGG AGC GTG AAC AA	163

Group name	Zip Code	C-zip code sequence (5'-3')	SEQ ID NO	Polymorphism position
SNOWELLA	*10*	GATCGGCCGGTGAAGCGAAAGGTTC	164	592-G
TOXIC ANABAENA	*12*	GATGGTGATCCCGCGCGTGCCGAAA	165	1293-G
APHANIZOMENON CLUSTER I	*14*	GGATTGCACCGTCAGCACCACCGAG	166	748-T
APHANIZOMENON CLUSTER II	*15*	TCCCAGGACGGCGCTGGCACGTTGA	167	1010-A
ANABAENA CLUSTER I	*16*	CGGCGTCCACGTCGAGTTCCTTCGC	168	658-T
ANABAENA CLUSTER II	*17*	TGTGCGCCCGAGATCGGTATCCCCG	169	635-G
ANABAENA+APHANIZOMENON SUBCLUSTER	*18*	ATCGCATCGTGATGGCGTAAGCTCC	170	1022-A

Fig. 39

METHODS FOR DETECTING TOXIC AND NON-TOXIC CYANOBACTERIA

[0001] This invention relates to a method for detecting toxic and non-toxic cyanobacteria. This invention relates also to oligonucleotides, which can be used in the detection method.

BACKGROUND OF THE INVENTION

[0002] Cyanobacteria produce a wide variety of bioactive compounds. Many of these are potent toxins, which cause health problems for animals and humans when producer organisms occur in masses in lakes and water reservoirs (Sivonen and Jones, 1999). Most well known of the cyanobacterial toxins are the hepatotoxic heptapeptides, microcystins. The general structure of microcystins is cyclo(-D-Ala-X-D-MeAsp-Z-Adda-D-Glu-Mdha-), where X and Z are variable L-amino acids, D-MeAsp is D-erythro-β-methylaspartic acid, Mdha is N-methyldehydroalanine and Adda is 3-amino-9-methoxy-2,6,8-trimethyl-10-phenyl-deca-4,6-dienoic acid. More than 65 structurally different microcystins are known (Sivonen and Jones, 1999). Most common variants have L-leucine and L-arginine in the positions of X and Z, respectively, and demethylated forms are also frequently found. Toxicity of microcystins is caused by the inhibition of protein phosphatases 1 and 2A (MacKintosh et al., 1990). The level of inhibition varies depending on the structure, but the Adda and D-Glu moieties, which are almost invariable in microcystins, are essential for the inhibition (Goldberg et al., 1995) and hence for the toxicity.

[0003] Microcystins have been found predominantly in cyanobacteria of three platonic, bloom-forming genera, *Anabaena, Microcystis* and *Planktothrix* (Sivonen and Jones, 1999). Not all members of these genera make microcystins and both toxic and non-toxic strains occur in the same species. Toxic and non-toxic strains of *Anabaena, Microcystis* or *Planktothrix* cannot be separated based on the classical morphological taxonomy or ribosomal gene sequencing (Lyra et al., 2001). On the other hand, one stain may produce different microcystins and also other peptides simultaneously (Sivonen et al., 1992; Fujii et al., 1996; Fastner et al., 2001.

[0004] Peptide synthetase genes were shown to be required for the synthesis of microcystins (Dittmann et al., 1997). Recently, the gene clusters encoding microcystin synthetase were sequenced and characterized from the unicellular Microcystis aeruginosa (Nishizawa et al., 2000; Tillet et al., 2000) and from the filamentous *Planktothrix agardhii* (Christiansen et al., 2003). It was demonstrated that the microcystins biosynthesis is a combination of peptide and polyketide synthesis (Nishizawa et al., 2000; Tillet et al., 2000).

[0005] The microcystin synthetase gene region spans about 55 kb, and includes genes for peptide synthetases (mcyA, -B, -C), polyketide synthases (mcyD), mixed peptide synthetase and polyketide synthases (mcyE, -G), and tailoring enzymes Tirett. et al. (2000), Nishizawa et al. (2000).

[0006] Microcystin producers among the filamentous, nitrogen-fixing genus, Anabaena, are found in North America, in France and in Northern Europe, where they frequently develop massive growth in lakes and reservoirs (Sivonen and Jones, 1999). The bioactive peptides produced by Anabaena 90 have been characterized: three microcystins (MCYST-LR, MCYST-RR and D-Asp-MCYST-LR; Sivonen et al., 1992), two seven-residue depsipeptides (anabaenopeptilide 90A and 90B), and three six-residue peptides having an ureido linkage (anabaenopeptins A, B and C; Fujii et al., 1996). However, the microcystin synthetase gene region from *Anabaena* has not been sequenced.

[0007] Based on the sequence data available, various DNA probes and primers have been designed and used to discriminate between toxic microcystin-producing and non-toxic non-microcystin producing genotypes by hybridization and PCR However, the existing primers deduced from *Microcystis* mcy genes, reliably identify potential microcystin-producers only in *Microcystis* and fail to amplify mcy sequences from part of microcystin containing strains of other genera. There is therefore a great need for oligonucleotides, which could be used as probes and primers in detecting toxic cyanobacteria also in genera other than *Microcystis*. Such oligonucleotides should discriminate between toxic microcystin-producing and non-toxic non-microcystin producing genotypes in various molecular biology methods, such oligonucleotides should be specific to the studied cyanobacteria genera and the oligonucleotides should be able to discriminate the most important or dominating microcystin producing cyanobacteria genera from one another.

[0008] It would be also of advantage if non-toxic cyanobacteria could be identified.

SUMMARY

[0009] It is the aim of the present invention to eliminate the problems associated with the prior art.

[0010] One object of this invention is to provide a method for the detection of toxic cyanobacteria.

[0011] In this invention it has been surprisingly found that by designing oligonucleotides to be specific for mcyE gene of the microcystin synthetase gene region, it is possible to detect cyanobacteria from all of the most potent toxin producing cyanobacteria genera. In addition it is possible to identify which cyanobacterial genus produces the toxin.

[0012] In particular, the oligonucleotides are designed to be specific for a region of mcyE gene responsible for adding Adda and D-glutamate to the immature synthesis product.

[0013] More specifically, the oligonucleotides are designed to be specific for a region of mcyE gene region catalyzing a peptide synthesis between Adda-D-glutamate and dehydroalanine and to the adenylating region. It is assumed that the step of adding Adda-D-glutamate-dipeptide is decisive for toxicity of the product. However, it is surprising that oligonucleotides designed to be specific for this region are genus specific and at the same time capable of identifying cyanobacteria from all other toxin-producing genera. Oligonucleotides of this invention can identify toxin producers at least among *Anabaena, Microcystis, Planktothrix, Nostoc* and *Nodularia* genera.

[0014] In this invention the whole microcystin synthetase gene region from *Anabaena* was sequenced. Before this invention it had not been possible to compare the sequences

of microcystin synthetase gene region from the main microcystin-producing cyanobacteria genera.

[0015] The oligonucleotides of this invention can be used in detecting toxin-producing cyanobacteria by using various molecular biology methods. Such methods are for example hybridization, PCR, reverse transcriptase PCR, QRT-PCR, LCR, LDR and minisequencing.

[0016] These methods can be combined with a microarray method. In a preferred detection method ligase detection reaction (LDR) is used together with a microarray method. Another preferred detection method is quantitative PCR (QRT-PCR).

[0017] Furthermore, the oligonucleotides of this invention can be used in detecting toxin-producing cyanobacteria together with a detection method using oligonucleotides designed to be specific for any other mcy gene, such as mcyA or mcyD gene.

[0018] One highly preferred embodiment of this invention is the use of the oligonucleotides of this invention together with oligonucleotides designed to be specific for 16S rRNA gene. Cyanobacterial genera can be identified based on the 16S rRNA gene. When oligonucleotides designed to be specific for mcyE (or some other mcy gene, such as mcyD) and for 16S rRNA gene are used together for example in the microarray method, it is possible to detect and identify both toxin- and non-toxin-producing genera. It is of great advantage that the oligonucleotides designed to be specific for mcyE and for 16S rRNA gene can be used under the same conditions. The LDR can be carried out under the same conditions and the hybridization in microarray on the same slide. This makes the monitoring of non-toxin cyanobacteria- and toxin-producing cyanobacteria technically easy and much more useful.

[0019] The detection method of the present invention can also be combined with a detection method measuring microcystin concentration, cell number, cell density or biomass. For example, mcyE copy number can be determined together with microcystin concentration and cell density and the main putative microcystin producers can be indicated.

[0020] One object of this invention are fragments of mcyE gene which are responsible for adding Adda and D-glutamate to the immature synthesis product in microcystin synthesis. In particular, the fragments are responsible for adding Adda-D-glutate dipeptide to dehydroalanine. Such fragments are or are located in the sequences selected from the group comprising SEQ ID NO. 1 to SEQ ID NO: 34 as shown in FIG. **19** A to H or comprising sequences SEQ ID NO: 35 to SEQ ID NO: 39 as shown in FIG. **15** A to C.

[0021] One object of this invention are furthermore oligonucleotides designed to be specific for any of the above mentioned fragments of mcyE gene, in particular for sequences selected from the group comprising SEQ ID NO. 1 to SEQ ID NO: 34 as shown in FIG. **19** A to H or sequences SEQ ID NO: 35 to SEQ ID NO: 39 as shown in FIG. **15** A to C or for fragments of said sequences.

[0022] Preferred oligonucleotides are primers mcyE-F2 (SEQ ID Nos: 64), AnamcyE-12R (SEQ ID NO: 65) and MicmcyE-R8 (SEQ ID NO:66) which can be used for example in amplifying target (or sample) nucleic acid by PCR.

[0023] Preferred oligonucleotides are discriminating probes of SEQ ID NO: 40 to SEQ ID NO: 45 and common probes of SEQ ID NO: 46 to SEQ ID NO: 51, which can be used for example in the ligase detection reaction.

[0024] One object of this invention is furthermore the mcyE gene from the *Anabaena* genus encoding the amino acid sequence of SEQ ID NO: 67 or a sequence having at least 80% identity, preferably 90%, more preferably 95% identity to said sequence, or a fragment of said sequence having polymorphic sites which make possible of designing oligonucleotides to be specific for the fragment.

[0025] One further object of this invention is mcyE gene from *Anabaena* genus having the nucleic acid sequence SEQ ID NO: 68 or a sequence having at least 80% identity, preferably 90%, more preferably 95% identity to said sequence, or a fragment of said sequence having polymorphic sites which make possible of designing oligonucleotides to be specific for the fragment.

[0026] One object of this invention is furthermore the mcyD gene from the *Anabaena* genus encoding the amino acid sequence of SEQ ID NO: 69 or a sequence having at least 80% identity, preferably 90%, more preferably 95% identity to said sequence, or a fragment of said sequence having polymorphic sites which make possible of designing oligonucleotides to be specific for the fragment.

[0027] One further object of this invention is mcyD gene from *Anabaena* genus having the nucleic acid sequence SEQ ID NO: 70 a sequence having at least 80% identity, preferably 90%, more preferably 95% identity to said sequence, or a fragment of said sequence having polymorphic sites which make possible of designing oligonucleotides to be specific for the fragment.

[0028] One object of this invention are fragments of mcyD gene. Such fragments are or are located in the sequences selected from the group comprising SEQ ID NO. 131 to SEQ ID NO: 149 as shown in FIG. **38** A to F.

[0029] One further object of this invention are oligonucleotides which can be used as discriminating probes and which are selected from the group comprising SEQ ID NO: 71 to SEQ ID NO: 90, and common probes which are selected from the group comprising SEQ ID NO: 91 to SEQ ID NO: 110. These primers and probes can be used for example in the ligation detection reaction.

[0030] Still further object of this invention is a kit for the detection of toxic cyanobacteria by the microarray method. The kit preferably comprises

[0031] discriminating and common probes designed to be specific for mcyE gene;

[0032] DNA or RNA zip and complementary zip codes assigned to be specific for selected cyanobacterial genera.

[0033] One still further object of this invention is a kit for detection of toxic cyanobacteria by hybridization. The kit preferably comprises

[0034] primers designed to be specific for the mcyE gene;

[0035] probes designed to be specific for selected cyanobacterial genera.

[0036] In the kit may be used alternatively or in addition probes and primers designed to be specific for mcyD gene or other mcy gene.

[0037] According to a highly preferred embodiment the kit comprises in addition to probes and primers designed to be specific for mcy gene (such as mcyE and/or mcyD) also probes and primers designed to be specific for 16 S rRNA gene.

[0038] Other features, aspects and advantages of the present invention will become apparent from the following description and appended claims.

BRIEF DESCRIPTION OF THE FIGURES

[0039] FIG. **1**. The microcystin synthetase gene cluster of *Anabaena* strain 90, biosynthetic model for the formation of microcystin-LR and the general structure of microcystins. The symbols for the domains are: A, adenylation; C, condensation; T, thiolation; NMT, N-methyltransferase; EP, epimerase; TE, thioesterase; KS, β-ketoacyl synthase; AT, acyltransferase; CM, C-methyltransferase; DH, dehydratase; KR, β-ketoacyl reductase; ACP, acyl carrier protein; AMT aminotransferase. Adda is 3-amino-9-methoxy-2,6,8-trimethyl-10-phenyldeca-4,6-dienoic acid, X and Z are variable amino acids. The arrows point to three methyl groups, which are putatively introduced by the C-methyltransferase domains. The way of cyclization of the microcystin precursor is shown with an arrow on the right of the picture.

[0040] FIG. **2**. A. Comparison of the putative C-methyltransferase domains in McyG, McyD and McyE of *Anabaena* 90 with three bacterial C-methyltranferase domains in the region of the conserved motifs:

[0041] 1. (VIL)(LV)(DE)(VI)G(GC)G(TP)G; 2. (PG)(QT-)(FYA)DA(IVY(FI)(CVL) and 3. LL(RK)PGG(RIL)(LI)(L-FIV)(IL) (Kagan and Clarke, 1994). EpoE is the polyketide synthase in epothilone biosynthesis of *Sorangium cellulosum* (AF217189), HMWP1 is the high-molecular-weight-protein in yersiniabactin biosynthesis coded by irp1 of *Yersinia enterocolitica* (Y12527) and ECUbiE is *Escherichia coli* C-methyltransferase, UbiE (P27851). Residues in bold letters (in the boxed areas) are identical to the consensus amino acids of the motifs. Amino acids (outside of the boxed areas), which are identical in at least five of the six sequences, are shaded.

[0042] B. Alignment of the aminotransferase domain of *Anabaena* 90 McyE, AmcyEamt, with other known aminotransferase domains and with two aminotransferases of *Escherichia coli*. McyEamt and PmcyEamt are from mcyE of *Microcystis aeruginosa* PCC7806 (AF183408) and of *Planktothrix agardhii* CYA126 (AJ441056), respectively. ItuAamt is from itrin synthetase of *Bacillus subtilis* RB14 (AB050629) and MycAamt from mycosubtilin synthetase of *Bacillus subtilis* ATCC6633 (AF184956). ECGSA is glutamate-1-semialdehyde aminotransferase (F90648) and EcArgD is ArgD, acetylornithine aminotransferase (P18335). The conserved pyridoxal-5'-phosphate-binding residues (Mehta et al., 1993), an aspartate and a lysine, are marked with the asterisks. Amino acids, which are the same in at least five of the seven proteins, are shaded.

[0043] FIG. **3**. Motif sequence alignments of (A) dehydratase (DH) and (B) ketoreductase (KR) domains of *Anabaena* 90 microcystin synthetase, AMCD-DH2, AMCD-DH3, AMCG-KR1, AMCG-KR2 and AMCD-KR3, with rifamycin synthase, RifE-DH10 and RifE-KR10 (*Amycolatopsis mediterranei*; AF040570), and rapamycin synthase, RapA-DH4, RapB-DH10, RapA-KR4 and RapB-KR10, (*Streptomyces hygroscopicus*; X86780). The conserved residues of (A) the active site motif $H(X)_3G(X)_4P$ (Aparicio et al., 1996) and of (B) the NAD cofactor binding site, $GXGXX(G/A)(X)_3(G/A)$, (Scrutton et al., 1990) are marked with asterisks. Amino acids which are invariant in all proteins, are in bold letters (A) and (B). The numbers of the domains refer to the module of the particular synthase.

[0044] FIG. **4**. Comparison of the motifs in acyltransferase (AT) domains of the microcystin synthetases with the consensus sequences of malonyl and methylmalonyl loading AT domains described by Ikeda et al. (1999). AT domains (AT1-AT4) are from *Anabaena* 90, AMcyG, AMcyD and AMcyE, from *Microcystis aeruginosa*, MMcyG, MMcyD and MmcyE (AF183408) and from *Planktothrix agardhii*, PMcyG, PMcyD and PmcyE (AJ441056). Bold letters indicate the amino acids, which are significantly specific to malonyl loading domains, and underlined, bold letters point out the residues, which are specific to methylmalonyl loading domains. Serines of the active site are marked with an asterisk.

[0045] FIG. **5**. Alignments of the β-ketoacyl synthase (KS) (A) and acyl carrier protein (ACP) (1B), domains of *Anabaena* 90 microcystin synthetase with the KS and ACP domains of rapamycin synthase, RapA-KS1, RapA-ACP1 and RapC-ACP11 (*Streptomyces hygroscopicus*, X86780) and of rifamycin synthase, RifA-KS1 and RifA-ACP1 (*Amycolatopsis mediterranei*, AF040570) near the active sites. (A) AMCG-KS, AMCD-KS1, AMCD-KS2 and AMCEKS are from the KS domains of *Anabaena* 90 McyG, McyD and McyE, respectively. An asterisk marks the active site cysteines. The identical amino acids are in bold letters. The two histidine residues, which are invariant in PKS and fatty acid synthases (Aparicio et al., 1996) are underlined. (B) AMCG-ACP, ACD-ACP1, AMCD-ACP2 are from the ACP domains of *Anabaena* 90 McyG, McyD and McyE. The active site motif, which frequently is LGxDS, is underlined. The serine residues, which bind phospho-pantetheine, are indicated by an asterisk.

[0046] FIG. **6**. The general structure of microcystins and nodularin. Microcystin is a cyclic peptide containing seven amino acids D-Ala-X-D-MeAsp-Z-Adda-D-Glu-Mdha, where X and Z represent variable L-amino acids, D-Me-Asp is D-erythro-β-methylaspartic acid, Mdha is N-methyldehydroalanine, and Adda is the β-amino acid, 3-amino-9-methoxy-2,6,8-trimethyl-10-phenyldeca-4,6-dienoic acid. Nodularin differs from microcystins by lacking the amino acids D-Ala and X, and having N-methyldehydrobutyrine Mdhb) in place of Mdha. The dashed line indicates the two amino acids absent in nodularins.

[0047] FIG. **7**. Congruence between the 16S rRNA and rpoC1 data set and the microcystin synthetase gene data set. (A) A maximum-likelihood tree based on the 16S rRNA and rpoC1 data set (−1 nL-8004.26493). Branch lengths are proportional to sequence change. Maximum likelihood and maximum parsimony bootstrap values from 1000 bootstrap replicates are given above and below the line respectively. (B) A maximum-likelihood tree based on the mcyA, mcyD and mcyE data set (−1 nL=8781.50660). Branch lengths are proportional to sequence change. Maximum likelihood and maximum parsimony bootstrap values from 1000 bootstrap replicates are given above and below the line respectively.

[0048] FIG. **8**. A maximum-likelihood tree based on the 16S rRNA gene showing the sporadic distribution of cyanobacterial genera known to produce microcystins. Strains of the genera *Planktothrix, Microcystis , Anabaena* and *Nostoc* produce microcystins while strains of the genus *Nodularia* produce nodularins. Toxic strains are indicated by bold font.

[0049] FIG. **9**. Cycle treshold (Ct) values obtained by microcystin synthetase E (mcyE) quantitative real-time PCR (QRT-PCR) with external A) *Anabaena* standard strains of *Anabaena* 90 (O), *Anabaena* 315 (□), and *Anabaena* 202A1 (Δ) as well as with B) those of *Microcystis* strains *Microcystis* GL 260735 (O), *Microcystis* PCC 7806 (□), and *Microcystis* PCC 7941 (Δ) as a function of mcyE copy numbers. Error bars, which are almost hidden by the symbols, give the standard deviation for three independent amplifications.

[0050] FIG. **10**. Microcystin concentration (x) (μg l^{-1}) determined with ELISA and *Anabaena* as well as *Microcystis* microcystin synthetase E (mcyE) copy numbers (copies ml^{-1}) obtained with quantitative real-time PCR using Lake Tuusulanjärvi water samples collected during summer 1999. Gene mcyE copy numbers were calculated with the external standards of *Anabaena* 202A1 (■), *Anabaena* 315 (□), *Microcystis* PCC 7806 (○) and *Microcystis* PCC 7941 (●).

[0051] FIG. **11**. Microcystin concentration (X) (μg l^{-1}) determined with ELISA and *Anabaena* as well as *Microcystis* microcystin synthetase E (mcyE) copy numbers (copies ml^{-1}) obtained with quantitative real-time PCR using lake water samples collected from different water depths of four Lake Hiidenvesi basins on 15 Aug. 2001. Gene mcyE copy numbers were calculated with the external standards of *Anabaena* 202A1 (■), *Anabaena* 315 (□), *Microcystis* PCC 7806 (○) and *Microcystis* PCC 7941 (●).

[0052] FIG. **12**. The cell numbers of the most dominant cyanobacterial genera in Lake Tuusulanjärvi in 1999 by light microscopy. The most dominant cyanobacterial genera were *Anabaena* (□), *Microcystis* (O) and *Aphanizomenon* (Δ).

[0053] FIG. **13**. The cell numbers of the most dominant cyanobacterial genera in Lake Hiidenvesi on 15 Aug. 2001 by light microscopy. The most dominant cyanobacterial genera were *Anabaena* (□), *Microcystis* (O) and *Aphanizomenon* (Δ. The samples were taken from different water depths at the four basins of Lake Hiidenvesi.

[0054] FIG. **14**. Clusters of group-specific mcyE gene consensus sequences.

[0055] FIG. **15**. A; B, C. 800 bp consensus sequence of mcyE from *Anabaena, Microcystis, Nodularia, Nostoc, Oscillatoria/Planktothrix* (SEQ ID NOs 35 to 39).

[0056] FIG. **16**. The principle of the DNA-chip (Microarray) method.

[0057] FIG. **17**. Deposition scheme of the mcyE probes. Deposition scheme obtained using a non-contact dispensing system. Each zip code was spotted ten times. The deposition quality of the Zip Code oligonucleotides on the slides has been checked by means of hybridisations with Cy3 labelled poly(dT) complementary to the $poly(dA)_{10}$ sequence of each Zip Code.

[0058] FIG. **18**. Hybridization results obtained using PCR amplified mcyE gene coming either from pure strains or from environmental samples as template in LDR.

[0059] FIG. **19** A-H. Alignment of 800 bp of nucleic acid sequences from 30 strains (+4 consensus sequences) from *Anabaena, Microcystis, Nodularia, Nostoc*, and *Oscillatoria/Planktothrix* genera (SEQ ID NOs 1 to 34).

[0060] FIG. **20**. List of polymorhism positions, group-specific probes (discriminating probes SEQ ID NOs 40 to 45 and common probes 46 to 51) and their correspondent Zip Codes and complementary Zip Codes SEQ D NOs 52 to 57 and 58 to 63.

[0061] FIG. **21**. Amino acid sequence encoded by *Anabaena* mcyE gene (SEQ ID NO 67).

[0062] FIG. **22** A-D. Nucleic acid sequence of *Anabaena* mcyE gene (SEQ ID NO 68).

[0063] FIG. **23** A, B. Amino acid sequence encoded by *Anabaena* mcyD gene (SEQ ID NO 69).

[0064] FIG. **24** A-D. Nucleic acid sequence of *Anabaena* mcyD gene (SEQ ID NO 70).

[0065] FIG. **25**A. The cyanobacterial phylogenetic tree constructed using the NJ algorithm, according to a central database of processed sequences. ARB cyanobacterial 16S rRNA gene database we used contained 281 sequences from public databases and 59 from this study.

[0066] FIG. **25**B. Updated ARB tree with Snowella sequences.

[0067] FIG. **25**C. Updated ARB tree with subclustering of *Anabaena* and *Aphanizomenon* groups.

[0068] FIG. **26**. Main features of LDR method coupled to a Universal Microarray.

[0069] Panel A: After the hybridization of a discriminating probe and a common probe to the target sequence (16s rRNA gene), ligation occurs only if there is perfect complementarity at the junction between the two probes. The reaction is thermally cycled.

[0070] Panel B: The LDR product is hybridized to an addressable Universal Microarray, where unique Zip code sequences have been spotted.

[0071] FIG. **27** A. Deposition scheme obtained using a contact dispensing system. Each Zip code was spotted four times, except universal Zip code (twelve times) and the Zip code corresponding to hybridization control (eight times). The deposition quality of the Zip Code oligonucleotides on the slides has been checked by means of hybridisations with Cy3 labelled poly(dT) complementary to the $poly(da)_{10}$ sequence of each Zip Code.

[0072] FIG. **27**B. Deposition scheme of Universal Array for the detection of toxic and non-toxic cyanobacteria. The Universal Array is made of 8 subarray per slide. Each subarray is made of 208 spots including zipcodes for hybridization control, cyanobacterial universal probes, 16S rRNA gene specific probe, mcyE specific probe and empty spot as a negative control. Each specific zip code for the recognition of cyanobacteria universal probe, 16S RNA gene probe and mcyE gene probe is spotted in quadruplicate. The LDR positive control (zipcode no 63) is replicated 6 times, while the hybridization positive control (zipcode no 66) is replicated 8 times.

[0073] FIG. **28**. Some results obtained using as LDR template PCR amplified 16S rRNA gene coming either from pure strains (both axenic and isolated in this study) or from cloned rDNA sequences.

[0074] Panel A: *Aphanizomenon* sp. 202; Panel B: Calothrix marchica Bai 71-96; Panel C: Leptolyngbya OBB19S12; Panel D: Lyngbya OBB32S04; Panel E: *Microcystis* 1BB 38S; Panel F: *Nodularin* sp. PCC73104/1; Panel G: Plankthotrix 1LT27S08; Panel H: *Spirulina subsalsa* PCC6313; Panel I: *Synechococcus* Heg 74-30; Panel J: Woronichinia OES46; Panel K: *Cylindrospermum stagnale* PCC7417; Panel L: *Synechocystis* PCC 6905; Panel M: *Nostoc* sp. 152; Panel N: *Anabaena*; Panel O: Cyanothece PCC 7418.

[0075] FIG. **29**. Hybridization results obtained using LDR artificial mixes with unbalanced amounts of PCR products derived from the following cyanobacterial samples: *Aphanizomenon* sp. 202, *Microcystis* OBB 34S, *Sprirulina subsalsa* PCC6313, *Calothrix* sp. PCC7714, Woronichinia OES46 clone. Different ratios have been used: 100:1, 50:1, 100:5, 50:5, in which *Aphanizomenon* sp. 202 and Microcystis OBB 34S have been the more concentrated samples.

[0076] Panel A: Unbalanced 100:1 LDR mix, Panel B: 50:1 LDR mix; Panel C: 100:5 LDR mix; Panel D: 50:5 LDR mix; Panel E: unbalanced LDR mix performed with 500 fmol of the amplicon derived from Microcystis OBB 34S and 5 fmol of the PCR fragment obtained from Woronichinia OES46 clone.

[0077] FIG. **30**A. Comparison of the results obtained using two LDR unbalanced mixes 100:1 (100 fmol of Microcystis OBB 34S and 1 fmol each of Spirulina, Woronichinia and Calothrix).

[0078] Panel A: The LDR unbalanced mix was prepared using 4U of Pfu DNA ligase.

[0079] Panel B: 8U of the enzyme was added-in the same LDR unbalanced mix described above.

[0080] FIG. **30**B. 16S and mcyE detection onto universal Array. Example of quantification.

[0081] FIG. **31**. Linear correlation between signal intensity and template concentration

[0082] FIG. **32**. List of the group-specific 16S rRNA gene probes and their correspondent Complementary zip codes (SEQ ID NOs 111 to 130) (discriminating probes SEQ ID NOs 71 to 90, common probes SEQ ID NOs 91 to 110).

[0083] FIG. **33**A, B. Cyanobacterial strains used to validate the LDR probes.

[0084] FIG. **34**. Clones of 16S rRNA gene libraries obtained from environmental samples and used in the LDR reaction.

[0085] FIG. **35**. PCR amplification from genomic DNA using 16S cyano primers and mcyE primers; primer F=mcyE-F2 and primer R=mcyE-R4; amplification protocol: 1×(3', 95° C.), 30×(30", 94° C.; 30", 56° C.; 1', 72° C.), 1×(10', 72° C.).

[0086] FIG. **36**. Ligation Detection Reaction for toxic and non-toxic cyanobacteria recognition.

[0087] FIG. **37**. Hybridization on DNA chip.

[0088] FIG. **38** A to F mcyD sequence fragments from different cyanobacteria genera (SEQ ID Nos 131-149). In SEQ ED Nos 137, 138 and 139 N is T.

[0089] FIG. **39**. List of the group-specific 16S rRNA gene probes (discriminating probes SEQ ID NOs 150 to 156) (common probes SEQ ID NOs 157 to 163) and C-zip Code sequences (SEQ ID Nos 164 to 170).

DETAILED DESCRIPTION OF THE INVENTION

Definitions

[0090] By "nucleic acid from a biological sample" is in this invention meant any target or sample nucleic acid, which originates from an environmental sample, such as water, soil cyanobacterial bloom, cyanobacterial culture, mixed population of cyanobacteria and other microbes etc. Nucleic acid is usually DNA, but in can be also RNA. The nucleic acid is usually extracted from the sample by conventional means known for the skilled artisan, but may also be liberated by repeated freeze-thawing to disrupt cellular integrity, or cells are used directly from the sample.

[0091] These techniques may also comprise the step of amplifying the nucleic acid before analysis. Amplification techniques are known to those of skill in the art and include, but are not limited to cloning, polymerase chain reaction (PCR), ligase chain reaction (LCR), nested polymerase chain reaction, self sustained sequence replication (Guatelli, J. C. et al., 1990, Proc. Natl. Acad. Sci. USA 87:1874-1878), transcriptional amplification system (Kwoh, D. Y. et al., 1989, Proc. Natl. Acad. Sci. USA 86:1173-1177), and Q-Beta Replicase (Lizardi, P. M. et al., 1988, Bio/Technology 6:1197).

[0092] The oligonucleotides of this invention are brought into contact with the target or sample nucleic acid under suitable conditions, which depend on the chosen molecular biology method, such as hybridization, PCR, LDR etc.

[0093] By "an oligonucleotide designed to be specific for the mcyE gene" it is meant that by using nucleic acid sequence data from several cyanobacterial genera and from several species of the genera, an oligonucleotide is designed to be specific for the mcyE gene of the microcystin synthetase operon. The length of an oligonucleotide may be 10 to 150 nucleotides depending on the detection method used. An oligonucleotide for hybridization is at least 20 bp, for PCR at least 10 bp and for LDR at least 15 bp.

[0094] Any probe or primer can be prepared according to methods well known in the art and described, e.g., in Sambrook, J. Fritsch, E. F., and Maniatis, T. (1989 (Molecular Cloning: A Laboratory Manual, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y. For example, discrete fragments of the DNA can be prepared and cloned using restriction enzymes. Alternatively, probes and primers can be prepared using the Polymerase Chain Reaction CPCR) using primers having an appropriate sequence.

[0095] Primers and probes (RNA, DNA) described herein may be labeled with any detectable reporter or signal moiety including, but not limited to radioisotopes, enzymes, antigens, antibodies, spectrophotometric reagents, chemiluminescent reagents, fluorescent and any other light producing chemicals. Additionally, these probes may be modified with-

out changing the substance of their purpose by terminal addition of nucleotides designed to incorporate restriction sites or other useful sequences.

[0096] These probes may also be modified by the addition of a capture moiety (including, but not limited to paramagnetic particles, biotin, fluorescein, dioxigenin, antigens, antibodies) or attached to the walls of microtiter trays to assist in the solid phase capture and purification of these probes and any DNA or RNA hybridized to these probes. Fluorescein may be used as a signal moiety as well as a capture moiety, the latter by interacting with an anti-fluorescein antibody.

[0097] By "a fragment of the mcyE gene" is meant principally any fragment of the mcyE gene which makes it possible to prepare oligonucleotides capable of identifying the mcyE gene from all of the microcystin producing genera and on the other hand is capable of discriminating different cyanobacterial genera from each other. The fragment is preferably related to the region of mcyE gene responsible for adding Adda and D-glutamate to the immature synthesis product. In particular, the fragment is related to the region catalyzing a peptide synthesis between Adda-D-glutamate and dehydroalanine and to the adenylating region. More specifically, the fragment is related to the region encoding the end part of the adenylation domain, the phospho-pantetheine binding site and the beginning of the domain which catalyses a peptide bond between D-glutamate and dehydroalanine. The length of the fragment may be between about 500 to 1000 nucleotides, which makes the alignment of nucleic acid sequence data from several cyanobacterial genera and species moderate to handle.

[0098] Examples of suitable fragments are the sequences of SEQ ID NO. 1 to SEQ ID NO: 34 as shown in FIG. **19** A to H or the consensus sequences SEQ ID NO: 35 to SEQ ID NO: 39 as shown in FIG. **15** A to C.

[0099] By "a fragment of the mcyD gene" is meant principally any fragment of the mcyD gene which makes it possible to prepare oligonucleotides capable of identifying the mcyD gene from all of the microcystin producing genera and on the other hand is capable of discriminating different cyanobacterial genera from each other.

[0100] Examples of suitable mcyD fragments are the sequences of SEQ ID NO. 131 to SEQ ID NO: 149 as shown in FIG. **38** A to F.

[0101] By "a suitable molecular biology method" is meant the chosen molecular biology method suitable for the purposes of detecting toxic cyanobacteria. The method may be selected from the group comprising hybridization, PCR, QRT-PCR, LCR, LDR and minisequencing.

[0102] PCR refers to the method for increasing the concentration of a segment of a target sequence in a mixture of genomic DNA without cloning or purification. This process for amplifying the target sequence consists of introducing a large excess of two oligonucleotide primers to the DNA mixture containing the desired target sequence, followed by a precise sequence of thermal cycling in the presence of a DNA polymerase. The two primers are complementary to their respective strands of the double stranded target sequence. To effect amplification, the mixture is denatured and the primers then annealed to their complementary sequences within the target molecule. Following annealing, the primers are extended with a polymerase so as to form a new pair of complementary strands. The steps of denaturation, primer annealing, and polymerase extension can be repeated many times (i.e., denaturation, annealing and extension constitute one "cycle"; there can be numerous "cycles") to obtain a high concentration of an amplified segment of the desired target sequence. The length of the amplified segment of the desired target sequence is determined by the relative positions of the primers with respect to each other, and therefore, this length is a controllable parameter. By virtue of the repeating aspect of the process, the method is referred to as the "polymerase chain reaction" (hereinafter "PCR"). Because the desired amplified segments of the target sequence become the predominant sequences (in terms of concentration) in the mixture, they are said to be "PCR amplified." In addition to genomic DNA, any oligonucleotide or polynucleotide sequence can be amplified with the appropriate set of primer molecules. In particular, the amplified segments created by the PCR process itself are, themselves, efficient templates for subsequent PCR amplifications. With PCR, it is possible to amplify a single copy of a specific target sequence in genomic DNA to a level detectable by the device and systems of the present invention.

[0103] PCR oligonucleotide primers or probes may be derived from either strand of the duplex DNA. The primers or probes may consist of the bases A, G, C, or T or analogs and they may be degenerated at one or more chosen nucleotide position(s). The primers or probes may be of any suitable length and may be selected anywhere within the DNA sequences from selected sequences which are suitable. In order to produce primers to a mcyE PCR, the mcyE gene(s) is typically examined using a computer algorithm, which starts at the 5' or at the 3' end of the nucleotide sequence. Typical algorithms will then identify oligomers in pairs of defined length that are unique to the gene, have a GC content within a range suitable for hybridization, and lack predicted secondary structure that may interfere with hybridization. The number of oligonucleotide pairs may range from two to one million.

[0104] Minisequencing reaction refers to a type of single base extension sequencing reaction using sequence terminators. In certain embodiments, minisequencing reactions are performed in the substantial absence of free single nucleotides, to minimize or prevent polymerization of nucleic acid beyond the single nucleotide sequenced by the sequence terminator. In certain embodiments, sequence terminators are labeled with fluorescent dyes, so that each nucleotide (A, G, T, or C) is identifiable by the color of the fluorescent label.

[0105] QRT-PCR or quantitative real-time PCR method involve measuring the amount of amplification product formed during an amplification process. Fluorogenic nuclease assays are one specific example of a real time quantitation method that can be used to detect and quantitate transcripts of present invention. In general such assays continuously measure PCR product accumulation using a dual-labeled fluorogenic oligonucleotide probe, an approach frequently referred to in the literature simply as the "TaqMan" method. The probe used in such assays is typically a short (ca. 20-25 bases) polynucleotide that is labeled with two different fluorescent dyes. The 5" terminus of the probe is typically attached to a reporter dye and the 3" terminus is

attached to a quenching dye, although the dyes can be attached at other locations on the probe as well. For measuring a mcyE transcript, the probe is designed to have at least substantial sequence complementarity with a probe binding site on a mcyE transcript. Upstream and downstream PCR primers that bind to regions that flank mcyE are also added to the reaction mixture for use in amplifying the mcyE polynucleotide. When the probe is intact, energy transfer between the two fluorophors occurs and the quencher quenches emission from the reporter. During the extension phase of PCR, the probe is cleaved by the 5" nuclease activity of a nucleic acid polymerase such as Taq polymerase, thereby releasing the reporter dye from the polynucleotide-quencher complex and resulting in an increase of reporter emission intensity that can be measured by an appropriate detection system.

[0106] Hybridization is used in reference to the pairing of complementary nucleic acids. Hybridization and the strength of hybridization (i.e., the strength of the association between the nucleic acids) is impacted by such factors as the degree of complementary between the nucleic acids, stringency of the conditions involved, the Tm of the formed hybrid, and the G:C ratio within the nucleic acids. For example stringent hybridization conditions are defined in Sambrook et al. 1989.

[0107] Ligation Detection Reaction LDR is based on the discriminative properties of the DNA ligation reaction. It requires the design of two probes specific for each target sequence, as described by Barany and co-workers (1999). One oligonucleotide brings a fluorescent label or other detection label and the other a unique sequence named complementary Zip Code (cZip Code). Ligated fragments, obtained in the presence of a proper template by the action of a DNA ligase, are addressed to the location on the microarray where the Zip Code sequence has been spotted. Such an array is therefore "Universal" being unrelated to a specific molecular analysis.

[0108] When two complementary pairs of probe elements are utilized, the process is referred to as the ligase chain reaction which achieves exponential amplification of target sequences (F. Barany, "The Ligase Chain Reaction (LCR) in a PCR World," PCR Methods and Applications, 1:5-16 (1991)).

[0109] As used herein "Arrays" or "Microarrays" refers to an array of distinct polynucleotides or oligonucleotides synthesized on a substrate, such as paper, nylon or other type of membrane, filter, chip, glass slide, or any other suitable solid support. The microarray can be prepared and used according to the methods described, for example in Lockhart, D. J. et al. (1996; Nat. Biotech. 14: 1675-1680) and Schena, M. et al. (1996; Proc. Natl. Acad. Sci. 93: 10614-10619).

[0110] The microarray or detection kit is preferably composed of a large number of unique, single-stranded nucleic acid sequences, usually either synthetic antisense oligonucleotides or fragments of cDNAs, fixed to a solid support. The oligonucleotides are preferably about 6-60 nucleotides in length, more preferably 15-30 nucleotides in length, and most preferably about 20-25 nucleotides in length. For a certain type of microarray or detection kit, it may be preferable to use oligonucleotides that are only 7-20 nucleotides in length The microarray or detection kit may contain oligonucleotides that cover the known 5', or 3', sequence, sequential oligonucleotides which cover the full length sequence; or unique oligonucleotides selected from particular areas along the length of the sequence. Polynucleotides used in the microarray or detection kit may be oligonucleotides that are specific to a moyE gene or genes of interest.

[0111] The nucleotide sequence data can be aligned and clustered according to their phylogenetic lineages so that "group-specific" consensus sequences are yielded: *Anabaena, Microcystis, Nodularia, Nostoc, Oscillatoria/Planktothrix*. Then, "group-specific" probes can be designed using a suitable database, such as ARB database named "robe design". Among the set of probes, discriminating probes with 3' position unique to each group in order to obtain ligase discrimination can be selected. After hybridization of a discriminating probe and a common probe to the target sequence, ligation occurs only if there is perfect complementarity at the junction between the two oligos. Common probes are designed immediately 3' to the discriminating oligo from the group-specific consensus and the detection is made by microarray method.

[0112] Zip code sequences can be selected randomly from those described by Chen and co-workers, 2000. Each Zip code is randomly assigned to a single cyanobacterial group. Each common probe is synthesized to have the complementary Zip code (cZip code) affixed to its 3' end.

[0113] Examples of discriminating probes are SEQ ID NO: 40 to SEQ ID NO: 45 and of common probes SEQ ID NO: 46 to SEQ ID NO: 51 designed to be specific for mcyE gene.

[0114] Examples of LDR zip codes are zip codes SEQ ID NO: 52 to SEQ ID NO:57.

[0115] Furthermore, examples of discriminating probes are SEQ ID NO: 71 to SEQ ID NO: 90 and of common probes SEQ ID NO: 91 to SEQ ID NO: 110 designed to be specific for 16S rRNA gene.

[0116] The method of the present invention can be used to detect toxic cyanobacteria at least from the genera *Anabaena, Microcystis, Planktothrix, Nostoc* and *Nodularia*.

[0117] The method can be combined if desired with a detection method using oligonucleotides designed specific for any other mcy genes or for 16S rRNA gene. A method based on 16S rRNA gene detection is in particular useful, if non-toxic cyanobacteria should be identified in addition to toxic cyanobacteria, when for example the condition of environment is monitored.

[0118] The method of this invention can be combined also with methods determining microcystin concentration, cell density, cell number, biomass, biovolume, chlorophyll-a, total RNA/DNA concentrations etc.

[0119] A kit for the detection of toxic cyanobacteria by microarray method preferably comprises

[0120] discriminating and common probes designed to be specific for the mcyE gene;

[0121] DNA or RNA zip and complementary zip codes assigned to be specific for certain cyanobacteria genera.

[0122] A kit for the detection of toxic cyanobacteria by hybridization preferably comprises

[0123] primers designed to be specific for the mcyE gene;

[0124] probes designed to be specific for certain cyanobacteria genera.

[0125] In the kit can be in addition to primers or probes designed to be specific for the mcyE and/or mcyD gene also primers or probes designed to be specific for 16 S rDNA.

[0126] In this invention we have identified and characterized the genes for the biosynthesis of hepatotoxins, microcystins from the filamentous, nitrogen fixing cyanobacterium *Anabaena* strain 90. Microcystin synthetase genes are now known from three different cyanobacterial genera, *Anabaena, Microcystis* and *Planktothrix*, which are the main producers of the microcystins. The arrangement of the genes is different between these species. The order of the domains, which are coded by two sets of the genes, is co-linear with the hypothetical sequence of the enzymatic reactions for microcystin biosynthesis only in *Anabaena* 90.

[0127] These genes provide extensive sequence information for the design of primers to be used in PCR-based methods for the sensitive detection, identification and quantification of producers of hepatotoxic microcystins and nodularins.

[0128] Identifying the most potent microcystin producer in a lake could be valuable knowledge e.g. in designing lake restoration strategies. In connection of this invention we identified the microcystin producing genera and quantified the microcystin synthetase gene E (mcyE) copy numbers in two lakes (Lake Tuusulanjärvi and Lake Hiidenvesi) by quantitative real-time PCR. Microcystin concentrations and cyanobacterial cell densities of these lakes were also determined. The main microcystin producer in Lake Tuusulanjärvi was *Microcystis* sp., since average Microcystis mcyE copy numbers were over 30 times more abundant than those of *Anabaena*. Lake Hiidenvesi seemed to contain both nontoxic and toxic *Anabaena* as well as toxic *Microcystis* strains. Microcystin concentrations of Lake Tuusulanjärvi and Lake Hiidenvesi correlated positively with *Microcystis* mcyE copy numbers.

[0129] mcyE sequences from *Anabaena, Microcystis, Nodularia, Nostoc* and *Oscillatoria/Planktothrix* were used for detecting polymorphic positions useful for detecting cyanobacterial strains using several different biomolecular techniques. These unique features were used for designing probes for cyanobacterial detection and identification by LDR in combination with a microarray.

[0130] The molecular classification of cyanobacteria is based on 16S rRNA gene sequences obtained from pure cultures (Wilmotte & Herdmann, 2001). Using this molecular information, several techniques can be used to determine the cyanobacterial composition of an environmental sample. The most widely used method is the 16S rRNA gene amplification with cyanobacterial specific PCR primers, cloning, sequencing and phylogenetic reconstruction (Giovannoni et al., 1988). This strategy is very time consuming and therefore is not suited to large scale screenings. Recently, DGGE and TGGE have been widely applied to molecular ecological research (Muyzer, 1999). However, the excision of bands, reamplification and sequencing are necessary to obtain a precise diversity analysis.

[0131] Oligonucleotide microarrays (microchips) have a major role in genomics and have gained wide attention in molecular diagnostics. Microarray technology has a great potential in environmental diagnostics. In fact, the DNA microarray technology has already been applied for microbial diversity detection. Microarrays have been used for quantitation of target microbial populations for environmental analysis (Guschin et al., 1997).

[0132] Rudi and coworkers (2000) designed a small cyanobacterial specific microarray for *Microcystis, Planktothrix, Anabaena, Aphanizomenon, Nostoc* and *Phormidiun.*

[0133] DNA microarray and the magnetic-capture hybridization technique have been combined to form a new technology named MAG-microarray. Bacterial magnetic particles (B3 MPs) on a MAG-microarray have been used for the identification of cyanobacterial DNA (Matsunaga et al., 2001). Genus-specific oligonucleotides probes for the detection of *Anabaena* spp., *Microcystis* spp., *Nostoc* spp., *Oscillatoria* spp. and *Synechococcus* spp. have been designed from the variable region of the cyanobacterial 16S rRNA gene of 148 strains. These probes have been immobilized on BMPs via streptavidin-biotin conjugation and employed for magnetic-capture hybridization against digoxigenin-labeled cyanobacterial 16SrRNA gene. Bacterial magnetic particles have been magnetically concentrated, spotted in a microwell on MAG-microarray and detected. The entire process of hybridization and detection has been automatically performed and all the five cyanobacterial genera have been successfully discriminated.

[0134] Recently, we have presented a Universal DNA Array approach to discriminate some groups of bacteria (Busti et al., 2002). This procedure, based on the discriminative properties of the DNA ligation reaction, requires the design of two probes specific for each target sequence, as described by Gerry and co-workers (1999). One oligonucleotide brings a fluorescent label and the other a unique sequence named complementary Zip Code (cZip Code). Ligated fragments, obtained in presence of a proper template by the action of a DNA ligase, are addressed to the location on the microarray where the Zip Code sequence has been spotted. Such an array is therefore "Universal" being unrelated to a specific molecular analysis.

[0135] Here we present the Universal DNA Array approach applied to the detection of cyanobacterial diversity. We designed probes specific for 19 different cyanobacterial groups (phylogenetic lineages including *Anabaena/Aphanizomenon, Calothrix, Cylindrospermopsis, Cylindrospermum, Gloeothece, Halotolerants, Leptolyngbya, Lyngbya, Microcystis, Nodularia, Nostoc, Oscillatoria/Planktothrix, Phormidium, Prochlorococcus, Spirulina, Synechococcus, Synechocystis, Trichodesmium, Woronichinia*) identified from the phylogenetic tree obtained from the ARB database constructed in this study.

[0136] 13 axenic strains from culture collection, 38 isolated culture strains and 44 clonal fragments recovered from environmental samples were used for validation purposes with excellent results demonstrating a high discriminative power. The proposed approach is extremely sensitive (down to 1 fmol of PCR amplified 16S gene region are detectable) allowing for the analysis of unbalanced environmental samples. LDR coupled to Universal Microarray performed on PCR samples containing 100:1 ratios of different amplicons yielded the correct identification of the starting strains.

This approach is therefore amenable to the analysis of complex environmental samples.

[0137] The Universal array was used for the detection of toxic and non-toxic cyanobacteria by using probes designed to detect both the 16 rRNA and mcyE genes. In the presence of the proper DNA template of both 16S rRNA and mcyE genes, the Universal Array functioned very well: only group specific spots, universal spots and the spots corresponding to the hybridization control showed positive.

Genes Coding for the Synthesis of Microcystins in *Anabaena*

The Order of the Genes in the Microcystin Synthetase Gene Cluster is Different in the Cyanobacterial Species

[0138] The arrangement of the genes is different in the gene clusters of microcystin biosynthesis from the strains of three species. In *Anabaena* strain 90, *Microcystis aeruginosa* (Tillett et al., is 2000; Nishizawa et al., 2000) and in *Planktothrix agardhii* CYA126 (Christiansen et al., 2003) the NRPS genes, mcyA, mcyB and mcyC have the same order, but the organization of the other genes is different. In *Anabaena* strain 90 and in *M. aeruginosa* the mcy-genes are in two clusters, which are transcribed in opposite directions, whereas in *P. agardhii* they are in one cluster transcribed in the same direction (except mcyT, which was not found in *Anabaena* and *Microcystis*). The arrangement of the genes from mcyD to mcyH in *Microcystis* is almost identical in *Planktothrix* (mcyF is missing in *Planktothrix*), but it differs from the order in *Anabaena*. In *Planktothrix*, compared to *Microcystis*, the part containing mcyD, mcyE, mcyF, mcyG, mcyH, mcyI and mcyJ is reversed. In this rearrangement, mcyF and mcyI were lost from the cluster and mcyJ was relocated after mcyG

The Biosynthesis of Microcystins

[0139] In *Anabaena*, the order of the domains coded by the genes in the two sets is co-linear with the hypothetical sequence of the enzymatic reactions for microcystin biosynthesis (FIG. **1**). The progression of the biosynthetic reactions follows the order of the functions coded first by mcyG and continuing with the activities coded by mcyD, mcyJ, mcyE, mcyF, mcyI, mcyA, mcyB and mcyC.

[0140] Phenyl acetate is the assumed starting unit in the biosynthesis of Adda (Moore et al., 1991). It is activated by the adenylating domain identified in the N-terminus of McyG, and transferred onto the subsequent thiolation (phosphopantetheine binding) site. Polyketide synthesis reactions are followed (FIG. **1**). All four extension units are malonyl-CoA molecules according to the substrate specificity of the AT domains (FIG. **4**). In McyG there is a KS domain to catalyse the first condensation reaction between phenylacetate and malonyl-CoA.

[0141] The reductive reactions needed to fashion the polyketide chain are putatively catalysed by KR and DH domains of McyD and McyE. The KR domain of McyG is in the right position to reduce the carbonyl group of the putative starter molecule. The methyltransferase domains of McyG, McyD and mcyE are the obvious candidates to introduce three methyl groups into the carbon frame of Adda. It was recently verified with a knockout mutant (Christiansen et al., 2003) that the incorporation of the fourth methyl, which is seen in the methoxy group of Adda, is catalysed by McyJ. The amino transferase domain of mcyE most likely adds the amino group, which participates in the peptide bond with the glutamate residue.

[0142] There are two condensation domains of peptide synthetases in McyE. The first one logically catalyses the peptide bond between Adda and glutamate, which is activated by the adenylation domain of McyE. The signature sequence, which was also determined as DPRHSGVVG for mcyE of both *M. aeruginosa* and *P. agardhii*, has no precedents in the databases (Table 2). The synthetases of other peptides, which contain glutamyl residues are known for bacitracin, fengycin and surfactin (accession numbers: AF007865, AF023464, AF087452 and D13262). In these compounds the standard α-carboxyl of glutamate is part of the peptide bond, while in microcystins it is the γ-carboxyl. This is analogous to the activation of aspartate/methylaspartate by the second adenylation domain of McyB. The β-carboxyl of aspartate/methylaspartate instead of the α-carboxyl is engaged in the peptide bond formation. This must have impact on the compositions of the glutamate and aspartate/methylaspartate binding pockets in the adenylation domains.

[0143] McyA has two adenylation domains for the activation of serine and alanine, respectively. The signature sequences of these domains have models and are almost identical in *Anabaena* 90, *M. aeruginosa* and *P. agardhii* (Table 2). The dehydration of serine supposedly takes place after the activation by adenylation and is catalysed by McyI, which is similar to phosphoglycerate dehydrogenases.

[0144] There is only one, internal, condensation domain in McyA, which most likely links dehydroserine and D-alanine. The bond between glutamate and dehydroserine is putatively catalysed by the C-terminal condensation domain of McyE. There is a methyltransferase domain in the first module of McyA for N-methylation of dehydroserine. The epimerase domain at the C-terminus of McyA converts L-alanine to the D-form.

[0145] Two modules of McyB and one module of McyC logically activate, and add three residues to the nascent peptide chain: L-leucine or L-arginine, methylaspartate or aspartate and L-arginine, respectively (FIG. **1**). The amino acids activated by the adenylation domains of McyC and by the first module of McyB (McyB-1) vary most frequently in microcystins. *M. aeruginosa* PCC7806 and *M. aeruginosa* K-139 produce mainly Mcyst-LR, and the substrate specificity conferring sequences in McyB-1 of these strains are identical with the signature sequence for leucine (Table 2). *M. aeruginosa* UV027 and *P. agardhii* CYA126 produce mostly Mcyst-RR, which is also produced by *Anabaena* 90 together with Mcyst-LR. Their signature sequences in McyB-1 are different and have no precedents in the databases (Table 2). In *M. aeruginosa* UV027 the specificity codes of McyB-1 and McyC are almost identical (DVWTIGAVE/DWTIGAVD) and match with the codes of McyC from *M. aeruginosa* K-139 and *M. aeruginosa* PCC7806, respectively (Table 2). Accordingly McyB-1 of *M. aeruginosa* UV027 and McyC activate arginine.

[0146] There is no epimerase domain in McyB of *Anabaena* 90 or in the other sequenced versions of McyB, though in microcystins, the aspartyl or methylaspartyl moiety is in the D-form. The epimerization in this position and in the glutamyl residue is putatively catalysed by McyF,

which in a BLAST search was similar to aspartate racemases, and was shown by Nishizawa et al., (2001) to complement a D-glutamate deficient mutant of *Escherichia coli*. The C-terminal thiosterase domain of McyC, as generally in nonribosomal peptide synthesis, (Kohli et al., 2001) catalyzes the final step in microcystin biosynthesis, the cyclization of the linear peptide (FIG. **1**).

[0147] McyH is probably not needed for the synthesis of microcystins but it may participate in the transport of microcystins.

[0148] In connection of this invention we obtained DNA sequences of three microcystin synthetase genes: mcyA, mcyE and mcyD. The mcyA gene fragment encodes part of the condensation domain, which catalyses a condensation reaction to form a peptide bond between the growing peptide and D-alanine. The fragment of the mcyE gene codes for a partial adenylation domain and a phospho-pantetheine-binding site, the region, which activates glutamic acid. The region of the mcyD gene encodes parts of both the β-ketoacyl synthase and the acyltransferase domains. We sampled representative producers of microcystins and nodularins (Table 1) in the genera *Anabaena, Microcystis, Planktothrix, Nostoc*, and *Nodularia*. Individual topologies generated from mcyA, mcyE and mcyD were rooted with homologues identified in BLAST searches. These topologies were congruent with one another (data not shown) and thus the data from all three genes were concatenated in order to increase the amount of information available in phylogenetic analyses.

Phylogenetic Evidence for the Early Evolution of Microcystin Synthesis

[0149] In order to investigate the role of horizontal gene transfer in the distribution of microcystin synthetase genes amongst cyanobacteria we assembled a data set comprised of 16S rRNA and rpoC1 sequences from the same set of taxa. These genes are conserved and widely used as tools for phylogenetic classification. No incongruence between the 16S rRNA and rpoC1 topologies could be found and the sequence data of these two genes was concatenated. We analysed these two data sets separately with maximum parsimony and maximum likelihood optimisation criteria. Bootstrap analyses were conducted to measure the stability of the observed phylogenetic patterns and revealed two well-supported topologies (FIG. **7**). The two maximum-likelihood topologies were perfectly congruent (FIG. **7**). The bootstrap support for the monophyly of the genera *Anabaena, Nodularia* and *Nostoc* was lower in the microcystin synthetase gene data set than in the 16S rRNA and rpoC1 data set (FIG. **7**). Likewise the bootstrap support for the monophyly of the genera *Planktothrix* and *Microcystis* was lower in the 16S rRNA and rpoC1 data set than in the microcystin gene data set (FIG. **7**). However, no conflicting nodes received bootstrap support above 45% in any analysis. Individual trees generated from mcyA (26 taxa), mcyE (30 taxa) and mcyD (19 taxa) all consistently supported the reciprocal monophyly of each genus (data not shown). In no instance was support for a lateral transfer recovered. The high degree of congruence between the microcystin synthetase gene data set and 16S rRNA and rpoC1 data set is consistent with an ancient origin of microcystins (FIG. **7**). This indicates that the phylogenetic marker genes and the microcystin synthetase genes have co-evolved for the entire length of the evolutionary history of this toxin. The sporadic distribution of microcystin synthetase genes in modern cyanobacteria suggests that the ability to produce the toxin has been lost repeatedly in the more derived lineages of cyanobacteria. Microcystins are one of the few known natural examples of combined polyketide synthase and peptide synthetase systems. Little is known about the evolution of these mixed polyketide and peptide synthetases and it is unclear whether the combination of these two systems is of recent origin. Congruence between the polyketide and peptide portions of the gene cluster as well as the 16S rRNA and rpoC1 data set demonstrates that the combination of these two systems is an ancient collaboration in the production of this toxin. Our results do not rule out the possibility that parts of the sequences of the microcystin synthetase gene cluster are of more recent origin. Indeed, the existence of many microcystin variants implies a fast evolution of certain gene domains.

[0150] Similarities in the chemical structures and biological action of microcystins and nodularins indicate that these compounds are closely related (Sivonen and Jones, 1999). However, the exact relationship between nodularins and microcystins remains ambiguous. Recent studies have suggested that the genes encoding microcystin synthetase have evolved from the genes encoding nodularin synthetase (Christiansen, 2003). Our data rejects the idea that nodularin synthesis predates microcystin synthesis (Christiansen, 2003 or that nodularin synthetase genes are a sister group to microcystin synthetases genes (Moffitt et al. 2001). Instead, our results suggest that nodularin synthetase genes are derived from microcystin synthetase genes and that nodularins should now be regarded as structural variants of microcystins. It is anticipated here that nodularin synthetase genes were formed from the ancestral microcystin synthetase gene set through a relatively recent deletion of the last mcyA module and the first mcyB module and by mutation changing the substrate specificity coded by the first module of mcyA. This finding is consistent with the production of nodularins by a single cyanobacterial genus and the limited structural variation of nodularins in comparison to microcystins Sivonen and Jones, 1999). Microcystins are commonly believed to have evolved in response to grazing pressure by zooplankton (DeMott et. al. 1991). Fossils of filamentous akinete-forming cyanobacteria are dated to 2000 million years ago (Amard et al., 1997).

[0151] This means that the *Anabaena, Nostoc*, and *Nodularia* genera and thus, the common ancestor of microcystin producing cyanobacteria are at least this old. Molecular clocks set a divergence time of 1576 million years ago for the crown eukaryotic lineages (Heckman, D. S. et al., 2001). Metazoans such as copepods and cladocerans are often envisaged as target organisms of microcystins (DeMott and Moxter, 1991). However, microcystin production predates all metazoans. If microcystins evolved as a chemical defense against zooplankton then the targets of the toxin must have been the early branching eukaryotes (Moon-van der Staay, S-Y. et al., 2001 and Brocks et al., 1999).

[0152] Protozoans are an underappreciated component of the zooplankton and may have been overlooked as the likely targets for the evolution of chemical defense in this case. It is not clear that microcystins evolved as a chemical defense and other proposed functions for microcystins include sid-

erophobic scavenging of trace metals such as iron (Utkilen and Gjolme, 1995) and a role in signalling and gene regulation (Dittmann et al, 2001).

[0153] Microcystins and nodularins are highly toxic to eukaryotic cells and pose a serious health risk to water users. Also the genera *Arthrospira* and *Aphanizomenon* are commonly used in health food supplements (Gilroy et al., 2000). Our study demonstrates that the ability to make microcystins has been lost repeatedly throughout the diversification of cyanobacteria. This means that toxin-producing strains may be found unexpectedly.

Quantification of Microcystin Synthetase E Copy Numbers of Microcystis and an *Anabaena* in Lakes by Quantitative Real-Time-PCR

[0154] In this invention a novel method to indicate the main putative microcystin producer of a lake is provided. The dominant putative microcystin producer was *Microcystis* in Lake Tuusulanjärvi and in the Basin of Kiihkelyksenselkä of Lake Hiidenvesi based on mcyE copy number quantification. This method enables to study in situ the responses of environmental factors on the growth of microcystin producing genera and could be used to observe the possible changes in cyanobacterial assemblages prior, during, and after lake restoration in order to find out, if the genus targeted lake restoration succeeded.

The Main Microcystin Producers

[0155] In Lake Tuusulanjärvi *Microcystis* spp. was the main putative microcystin producer, since average *Microcystis* mcyE copy numbers were clearly higher than those of *Anabaena* and thus, this result was in agreement with the higher cell numbers of *Microcystis* observed compared to those of *Anabaena*. Microcystin concentrations or hepatotoxicities have also presiously correlated positively with *Microcystis* spp. biomass in Lake Tuusulanjärvi (Ekman-Ekebom et al. 1992, Lahti et al. 1997). *Microcystis* spp. were also the main putative microcystin producers in the Basin of Kiihkelyksenselkä of Lake Hiidenvesi, although *Anabaena* cell numbers were higher than those of *Microcystis*. This indicates that majority of the *Anabaena* cells were nontoxic and Microcystis cells toxic in this basin. In the Basins of Mustionselkä, Nummelanselkä and Kirkkojärvi of Lake Hiidenvesi the main microcystin producer could not be assessed, since in the Basins of Mustionselkä and Nummelanselkä, the *Anabaena* and *Microcystis* mcyE copy numbers were quite similar and in the Basin of Kirkkojärvi the *Anabaena* and *Microcystis* mcyE copy numbers were below the detection limit. The low mcyE copy numbers detected in Kirkkojävi were in agreement with the low microcystin concentrations measured from this basin. Microcystin concentration correlated positively with *Microcystis* mcyE copy numbers with all studied samples whereas no significant correlation was found between microcystin concentrations and *Microcystis* and *Anabaena* cell numbers with all studied samples. Therefore, with microscope analysis it is not possible to determine reliably the most potent microcystin producer of a lake. Gene mcyE copy numbers, microcystin concentrations, and cyanobacterial cell densities were lower in Lake Hiidenvesi than in Lake Tuusulanjärvi. In Lake Tuusulanjärvi and in surface water of the Basins Nummelanselkä and Kiihkelyksenselkä of Lake Hiidenvesi WHO microcystin concentration guideline value for drinking water quality, 1 μg 1^{-1}, (Falconer et al., 1999.) was exceeded.

[0156] *Microcystis* and *Anabaena* mcyE copy numbers were one to over 200 times higher than the cell numbers observed with microscopy in Lake Tuusulanjärvi and Lake Hiidenvesi. In Lake Tuusulanjärvi Microcystis mcyE copy numbers increased after August in contrast to the cell density, which decreased. The explanation could be that after August cells had more genome copies or that the DNA of the lysed cells was present in the lake water and followed through the cell concentration and DNA extraction processes to the final DNA sample. Additional explanations for the high mcyE copy number and cell density ratio might be that the cell numbers detected with microscope were too low or the genome sizes of the external standard strains were underestimated. Even with the knowledge that cyanobacteria may have several genome copies in a cell (Becker, et al. 2002, Herdman et al., 1979, Labarre et al., 1989), it seems that the obtained mcyE copy numbers were too high. The genome sizes estimated for the *Anabaena* standard strains were 5.15 Mb according to the published data of *Anabaena* PCC 6309 and PCC 7122 (Castenholz, 2001). These *Anabaena* strains are nontoxic (Lyra, et al. 2001) and lack the microcystin synthetase genes, the sizes of which are not more than 53 or 55 kb (Christiansen et al., 2003, Nishizawa et al., 2000 and Nishizawa et al. 1999 and Trlnett et al. 2000 and Example 1). For *Microcystis* standard strains the genome size of 4.70 Mb was used according to the genome size of one of the external standard strains, *Microcystis* PCC 7941 (Castenholz, 2001).

[0157] In general, nontoxic strains do not contain mcy genes (Neilan et al., 1999 and Tillett et al. 2001). However, some strains may have fragments of microcystin synthetase genes or mutations within these genes (Kaebernick et al. 2001, Neilan et al. 1999 and Tillett et al. 2001). These strains can be amplified with may primers, although they are not able to produce toxins. However, the significant positive correlation between *Microcystis* mcyE copy numbers and microcystin concentration indicated that such nontoxic strains were probably not present in Lake Tuusulanjärvi and in Lake Hiidenvesi.

[0158] Amplification efficiency. *Microcystis* mcyE QRT-PCR amplification efficiencies with Lake Tuusulanjärvi water samples (0.78-0.99) were similar to those of *Microcystis* standards (0.86-0.94) and those of *Anabaena* standards (0.96-0.99), which is a prerequisite for correct mcyE copy number quantification of the lake water samples. These similar QRT-PCR amplification efficiencies also ensured that no PCR-inhibiting contaminants were present in the Lake Tuusulanjärvi DNA samples. However, *Anabaena* mcyE QRT-PCR amplification efficiencies with Lake Tuusulanjärvi water samples were higher than one. This result can be explained by competition for primer annealing sites between primers and homologous sequences (Becker et al. 2000, Suzuki et al. 1996, Wawrik et al. 2002) and this competition may lead to suppression of the target DNA (Suzuki et al. 1996). This phenomenon has been shown to occur not only in conventional PCR (Suzuki et al. 1996) but also in QRT-PCR (Becker et al. 2000, Wawrik et al. 2002), although quantification is achieved during the early logarithmic phase of the amplification (Heid et al., 1996). *Anabaena* and *Microcystis* mcyE sequences are homologous (Example 2). Since in Lake Tuusulanjärvi the concentration of competing *Microcystis* mcyE genes was higher than that of *Anabaena* mcyE genes, it is possible that the *Anabaena* mcyE copy numbers were underestimated. In addition, the

mcyE-F2 forward primer amplified *Anabaena* as well as *Microcystis* sequences and increased the amount of competing homologous sequences.

[0159] Detection range of mcyE copy number quantification. The mcyE QRT-PCR amplification was log-linear in a range of three to four orders of magnitude. With high DNA template concentration, 6.6×10^6 mcyE copies in a reaction, amplification was inhibited with the DNAs of *Anabaena* 90, *Anabaena* 202A1, *Microcystis* GL 260735, and *Microcystis* PCC 7941 strains, since obtained Ct values were lower than they should have been according to the regression equation or Ct values could not be detected at all. The inlubition was probably caused by contaminants that co-extracted with DNA during the DNA extraction and purification as shown previously (Wintzingerode et al. 1997). The lowest detection limit of *Anabaena* and *Microcystis* mcyE QRT-PCR amplification was 660 mcyE copies in a reaction. The error of the Ct values in QRT-PCR has been shown to be higher with low DNA template concentrations than with high template concentrations (Grüntzig et al. 2001). However, in this study the lowest mcyE copy number concentrations of the external standards had the same CV % as the other concentrations, 0.1-3.6%.

[0160] The utilization of the mcyE copy number results. In this study, putative microcystin producing *Anabaena* and *Microcystis* were detected in both studied lakes. In Lake Tuusulanjärvi and in the Basin of Kiihkelyksenselka of Lake Hildenvesi the dominant putative microcystin producer was *Microcystis* based on mcyE quantification. Reduction of nutrient loading and resuspension (Boers et al. 1991, Chorus and 1999, Reynolds, 1997) could be successfiul strategies to decrease the density of *Microcystis*, since these may decrease nitrogen as well as phosphorus concentrations of the water. In addition, lower nutrient concentrations could favor the growth of nontoxic *Microcystis* strains instead of toxic, since the biomass of nontoxic *Microcystis* strains has been demonstrated to be higher than that of toxic strains with low nutrient concentrations at the end of a laboratory experiment (Vezie et al. 2002). Lake Hiidenvesi seemed to have nontoxic and toxic *Anabaena* strains as well as toxic *Microcystis* strains. However, mcyE copy numbers should be monitored during the whole growth period in order to have a better understanding of the population dynamics of this lake. A reduction of the external phosphorus loading could affect the mass occurrences of nitrogen-fixing cyanobacteria negatively. It is however not known how the reduction of nitrogen fixlng-cyanobacteria would affect the growth of toxic *Microcystis* strains. At least, the presence of toxic *Microcystis* strains should be taken into account in land use management of the catchment area of Lake Hiidenvesi.

Oligonucleotides for Detection and Identification of Toxic Cyanobacteria

[0161] In this invention was developed the identification on mcyE gene region of polymorphisms specific for different toxic cyanobacterial groups identified from the phylogenetic tree obtained from 34 toxic cyanobacterial sequences. The polymorphic positions were used for designing probes for PCR, hybridization, primer extension, ligation and LDR. Probes for ligation have been used in combination with randomly chosen tag sequences appended 5' to the so called common primers in order to be used in the universal array approach. Validation against different samples demonstrate the robustness of the proposed polymorphisn and probes.

Molecular Analysis of Cyanobacterial Diversity by Microarrays on "PCR-Amplified" 16SrRNA gene

[0162] We aimed at designing and testing a microarray based system for cyanobacterial diversity identification. We selected a molecular strategy based on the amplification of the 16S rRNA gene region using cyanobacteria specific primers (Edwards et al. 1989, Lepre et al. 2000) followed by group discrimination based on a multiplexed ligation detection reaction performed employing proper probes. Ligated fragments characteristics of each group were demultiplexed on a Universal array. This approach, originally proposed by Gerry et al (1999) has found several application. We used the ARB database including 281 public sequences belonging to the 19 phylogenetic lineages we decided to target (*Anabaena/Aphanizomenon, Calothrix, Cylindrospermopsis, Cylindrospermum, Gloeothece, Halotolerants, Leptolyngbya, Lyngbya, Microcystis, Nodularia, Nostoc, Oscillatoria/Planktothrix, Phormidium, Prochlorococcus, Spirulina, Synechococcus, Synechocystis, Tnichodesmium, Woronichinia*). Not all of these groups are present in the environmental samples from the lakes involved in the MIDI-CHIP project but all them were included in order to allow for future research studies. Sequences were clustered as shown in FIG. **25**. For each group we calculated a consensus sequence with a cutoff of 75%. The resulting consensuses were aligned and group specific probes were searched along the entire 16S rRNA gene region. Following the LDR approach (FIG. **26**) we identified two unique probes for every group (a common probe and a discriminating probe). Selected probes were tested against the set of sequences of the corresponding group in order to verify the perfect match, in particular around the site of ligation. Then probe sequences were tested against the remaining cyanobacterial sequences in order to verify their selectivity. Selected probes are spread all over the entire 16S amplicon. Selected common probes were then randomly combined to a set of cZipCode sequences previously proposed for the Universal array approach (Gerry et al 1999, Chen et al 2000). Potential cross hybridization was checked by BLAST analysis of each common and discriminating probe against all others. Probes were then synthesized, HPLC purified and tested by mass spectrometry. This stringent quality assurance procedure is mandatory to achieve expected results in LDR. Ordinary PCR quality probes yielded poor performance due to low phosphorilation or Cy3 labeling and exceedingly high failure sequences. Similar quality controls were performed on the 5' amino-modified ZipCode sequences spotted by contact printing on Codelink Slides. We generated 8 subarrays per slide (96 spots per subarray including zipcodes for a hybridization control (eight spots at corners), cyanobacterial universal probes (12 spots in the middle and at corners) and 19 lineage-specific ZipCodes spotted in quadruplicate. Slides were batch-tested by hybridization using a labeled polyT probe matching the polyA tail appended in 5' to every ZipCode probe. In order to validate the designed probes we run a blank (no template) LDR. No signals were detected demonstrating that no false ligation occurred (this problem is often encountered when performing minisequencing (Lindroos, 2002). Then 51 strains of known 16S rRNAsequence belonging to 13 phylogenetic groups (FIG. **33**) were used to test the proposed system. FIG. **28** clearly illustrates LDR

specificity when using 100 fmol of each single template independently reacted against the complete set of probes. Six out of 19 groups were not included in the test panel due to their unavailability but their corresponding LDR probes were present in the LDR mix and did not generate any false positive result. It should be noted that, although not identical, the LDR/Universal array efficiency was very similar among all probes. Comparing the intensity between the cyanobacterial universal probe and each lineage specific probe, we found a ratio very close to 1 for most groups. (Here a graph showing this comparison could be more clear that the following-description). Probes for *Lyngbya, Nodularia, Anabaena* and *Cyanotizece* (FIG. **28** D, F, N, O respectively) consistently yielded higher efficiency. However the similarity of results using very different sequences having very close thermodynamic properties is a distinctive feature of this approach. Hybridization based arrays Loy, 2002; Rudy K. 2000) depend heavily on local sequence characteristics. When hybridization is performed in high salt buffers in a single stringency condition, large variability in signal intensity can be expected (Loy, 2002). On the contrary, using the exquisite sequence specificity of the ligation reaction (Gerry, 1999) and the very high annealing temperatures required during cycling, a very homogeneous behaviour is found. Very little influence of the sequence context has been demonstrated. Our results in a different sequence context, the highly polymorphic HLA region (Consolandi, 2003) further confirm these findings. Another distinctive feature of the LDR approach is related to the excellent sensitivity gained by means of a cycling procedure based on thermostable ligases. We were able to detect down to 1 fmol (around 2 ng) of PCR amplified material thanks to the linear amplification gained through LDR. FIG. **31** show the results we obtained using a serial dilution of Planktothrix 16S amplicon from 100 fmol to 1. A good linear relationship was found plotting the signal intensity against the concentration in a log scale.

[0163] The Universal array was used for the detection of toxic and non-toxic cyanobacteria designed to detect both the 16 rRNA and mcyE gene ligated probes. The ligation detection reaction was carried out under the same conditions by using an oligo mix containing both the probes for 16S rRNA gene and the probes for the mcyE gene. Finally the hybridization was carried on the same Universal Array where the 16S rRNA LDR product and, mcyE LDR product were detected.

EXAMPLES

Example 1

Genes Coding for the Synthesis of Hepatotoxic Heptapeptides (Microcystins) in the Cyanobacterium *Anabaena* Strain 90

Bacterial Strains and Culture Conditions

[0164] The cyanobacterial strain *Anabaena* 90 was isolated from Lake Vesijärvi, Finland and purified axenic (Sivonen et al., 1992; Rouhiainen et al., 1995). It was shown to produce three microcystins (MCYST-LR, MCYST-RR and D-Asp-MCYST-LR (Sivonen et al., 1992). *Anabaena* strain 90 was grown in Z8 medium (Kotai, 1972) without nitrate at ~22° C. with continuous illumination of 20-25 μmol $m^{-2}s^{-1}$. *Escherichia coli* strain DH5 α, which was used as a host for DNA cloning and sequencing, was cultured in Luria Broth at 37° C.

DNA Manipulations, Sequencing, Screening and Mapping of Cosmids

[0165] Extraction of cyanobacterial DNA and the preparation of genomic library has been described earlier (Rouhiainen et al. 2000). The genomic library was screened by colony hybrdization (Sambrook et al., 1989). The probe labelled with [^{32}P]dCTP was a 2.5 kb fragment from mcyA of *Microcystis aeruginosa* provided by Dr. Elke Dittmann (Humboldt University, Berlin). A total of about 6,000 colonies were tested. The insert DNA of 29 positive cosmid clones was mapped with HindIII, EcoRI and SpeI. The ends of 18 inserts were sequenced with SP6 and T7 primers, and the cosmid clones for sequencing the microcystin synthetase genes were selected. DNA of the cosmid clones was digested with restriction enzymes BstEII, HindIII, EcoRI, ScaI, SpeI or XbaI and ligated to pBluescript SK(+). Nested deletions and other DNA manipulations were performed according to Sambrook et al., (1989). Sequencing was carried out mainly by the University of Chicago Cancer Research Center DNA Sequencing Facility. Gaps were filled by amplifying chromosondal DNA in PCR with DyNAzyme™ EXT Polymerase (Finnzymes), the sequencing reactions were done with the BigDye Terminator Cycle Sequencing Kit (Applied Biosystems) and analyzed on the ABI 310 Genetic Analyzer. The standard T3 and T7 primers and oligonucleotides derived from already determined sequences were employed.

Sequence Analysis

[0166] Analysis and comparisons of sequences were performed with the Sequence analysis software package, version 8.0, University of Wisconsin Genetics Computer Group and with EMBOSS (European Molecular Biology Open Software Suite). CAP program (http://bioweb.pasteur.fr/seganal/interfaces/cap.html) was used for sequence assembly. Sequence similarity searches in databases were done with BLAST through the website of the National Center for Biotechnology Information http://www.ncbi.nlmnih.gov/BLASI). Searches for conserved domains and motifs were accomplished with the CD-Search program (http://www.ncbi.nlm.nih.gov/Structure/cdd/cdd.sht and with the Motif Scan program (http://hits.isb-sib.ch/cgi-bin/PFSCAN?). Clustal W was applied for multiple sequence alignments (http://npsa-pbil.ibcp.fr/cgi-bin/npsa_automat-.pl?page=npsa_clustalw.html).

Organization of the Microcystin Synthetase Genes

[0167] Microcystin synthetase genes in *Anabaena* strain 90 (mcyA-J) are organized in three putative operons (FIG. **1**) with a total size of 55.4 Kb. The first operon (mcyA-mcyB-mcyC) is transcribed in the opposite direction compared to the second (mcyG-mcyD-mcyJ-mcyE-mcyF-mcyI) and the third operon (mcyH). The ORFs mcyA and mcyG are separated by 1275 bp; mcyI and mcyH by 297 bp (FIG. **1**). The putative promoter regions were identified in front of mcyA (the −10 sequence, TAAATT, 315 bp and the −35

sequence, TTGTAT, 339 bp upstream from the translation start codon, ATG, of mcyA) and in front of mcyG (the −10 sequence, TATAAG, 145 or 223 bp and the −35 sequence, TTGACA, 172 or 250 bp upstream from the potential translation starts of mcyG). The promoter region was also identified before mcyH (the −10 sequence, TATAAA, 57 or 216 bp and the −35 sequence, TTGATA, 79 or 238 bp from the suggested translation initiation codons). Transcriptional starts prior to mcyD (distance 93 bp from mcyG), mcyE (37 or 95 bp from mcyJ), mcyF (42 bp from mcyE) and before mcyI (51 bp from mcyF) cannot be ruled out, although no transcription stop loops were identified following the preceding genes, and no Pribnow box could be identified in front of mcyD.

[0168] Characterization of the Peptide Synthetase Genes

[0169] In the first operon there are three open reading frames (ORFs) named mcyA, mcyB, and mcy. We suggest that the translation of mcyA starts with the ATG codon preceded (3 bp) by a potential ribosome binding site (RBS) GGAGAAG. The next ORF, mcyB, begins with an ATG codon 18 bp downstream from the previous stop codon (TAA) and 12 bp from a potential RBS AGAGGA. mcyC is overlapped by mcyB with one base pair. A putative RBS (ACGACAAG) is found 5 bp before the start codon ATG of mcyC. The lengths of mcyA, mcyB, and mcyC are 8364, 6399 and 3852 bp and they encode polypeptides with predicted masses of 315,663, 243,072, and 146,877 Da, respectively. The sequence analysis of mcA, mcyB, and mcyC revealed a typical modular structure for nonribosomal peptide synthetase (NRPS) genes (Marahiel et al., 1997) (FIG. **1**). mcyA contains two putative adenylation and thiolation domains, a condensation, an N-methyltransferase, and an epimerization domain. In mcyB there are two modules, both include condensation, adenylation, and thiolation domains. mcyC is composed of one module, containing a condensation, an adenylation, a thiolation, and a thioesterase domain (FIG. **1**).

Identification of the Polyketide Synthase Genes

[0170] The second operon contains six ORFs named mcyG-mcyD-mcyJ-mcyE-mcyF-mcyL A suggested translation start codon (ATG) of mcyG is located 8 bp downstream of a probable RBS (ACAGGA) giving an ORF (7827 bp), which could code for a protein of 2609 amino acids with a predicted mass of 289,859 Da. Another possible initiation is at an ATG, 75 bp upstream from the previously proposed start and 5 bp after a putative RBS (AAGGCA). This ORF. (7905 bp) possibly encodes a protein of 2635 amino acids, 292,851 Da. The ORFs mcyG and mcyD are separated by 96 bp. The translation of mcyD starts probably at an ATG codon 6 bp after a potential RBS (GGAAGGAG), consequently the size of this large ORF is 11,607 bp, encoding 3869 amino acids. Following the stop codon TAG of mcyJ there are 36 bp prior to a presumed ATG initiation codon of mcyE, which is preceded (5 bp) by a possible RBS (GCGGACAA). An alternative ATG start codon for mcyE is 57 bp downstream from the previously proposed one and 3 bp from a possible RBS (AATGGAGG). The two versions (10,446 bp and 10,386 bp) of this large ORF, mcyE, could code for polypeptides of 3482 amino acids, 388,755 Da and 3462 amino acids, 386,501 Da, respectively. The ORF mcyD encodes a polypeptide of 3869 amino acids with the predicted mass of 430,216 Da. mcyD was identified as a polyketide synthase (PKS) gene, whereas mcyG and mcyE have a combined NRPS/PKS gene structure (FIG. **1**).

The Additional Genes

[0171] We suggest that the ORF mcyJ is initiated with a GTG codon 59 bp downstream of the stop codon (TAA) of mcyD, and 5 bp from a putative Shine-Dalgarno sequence AGGAGAG. There is no ATG codon located nearby. Accordingly, mcyJ is predicted to be 930 bp in length.

[0172] A small ORF, mcyF, (756 bp), following mcyE, begins with an ATG codon 42 bp after the previous stop codon TAG and 6 bp from a putative RBS (GGAGAA). The distance between mcyF and the next ORF, mcyI, (1011 bp) is 54 bp, and an alleged RBS (AAGGTTAA) is found 6 bp upstream from the designated start codon ATG of mcyI. Downstream (295 bp) from the stop codon (TAA) of mcyI an ORF, mcyH, (1776 bp) was found. It presumably is initiated from the ATG codon 6 bp after a potential RBS (AAGATG). Another possible translation start codon (ATG) is found 159 bp downstream from the former one and 4 bp from a putative RBS (AGGCATGG). The sizes of these potential McyH polypeptides of 592 and 539 amino acids are 67,731 Da and 61,754 Da, respectively. mcyJ, mcyF and mcyI encode polypeptides of 310,252 and 337 amino acids with predicted masses of 35,812, 28,426, and 36,750 Da, respectively. McyF is similar to aspartate racemases, McyJ belongs to methyltransferases, and McyI is related to D-3-phosphoglycerate dehydrogenases. McyH contains a membrane spanning and an ATP-binding domain of ABC transporters. A BLAST search of McyH found 75% identity (in 589 aa) to NosG from *Nostoc* sp. GSV224 (AF204805) and 39% identity (in 543 aa) to the hypothetical ABC transporter ATP-binding protein SLL0182 of *Synechocystis* sp. PCC 6803 (Q55774).

Comparison of Microcystin Synthetase Genes

[0173] The microcystin synthetase genes were previously sequenced from *M. aeruginosa* strains PCC7806 (mcyA-mcyJ, Tillett et al., 2000), K-139 (mcyA-mcyI, Nishizawa et al., 2000) and UV027 (mcyA-mcyC, Raps et al., unpublished, accession no. AF458094), and from *Planktothrix agardhii* CYA126 (Christiansen et al., 2002). When *Anabaena* 90 sequences were compared to *M. aeruginosa* sequences, they revealed 65 to 75 (mcyJ 80%) percent identities at the amino acid level and 69 to 75 (mcyJ 79%) percent identities at the nucleotide level (Table 1). The arrangement of the microcystin synthetase genes from mcyD to mcyJ in *Anabaena* 90 is different from the organization in *M. aeruginosa* PCC7806, in *M. aeruginosa* K-139 (known from mcyD to mcyI) and in *Planktothrix agardhii* CYA126.

TABLE 1

Percentage identity of the microcystin synthetase genes/polypeptides from *Anabaena* strain 90 with the genes/polypeptides sequenced from other cyanobacteria and the mol % G + C of the genes.

	mcy/Mcy[a]									
	A	B	C	D	E	F	G	H	I	J
M. aeruginosa PCC7806	69/68	72/69	74/73	72/69	75/74	71/65	74/71	74/70	74/71	79/80
mol % G + C	41	39	37	40	39	38	38	35	40	39
M. aeruginosa K-139	69/68	71/69	74/73	72/69	75/75	71/65	74/71	74/70	74/72	
mol % G + C	41	39	37	40	39	37	38	36	39	
M. aeruginosa UV027	69/68	73/71	74/73							
mol % G + C	41	39	37							
P. agardhii CYA126/8	67/66	72/70	80/79	77/73	78/77		77/74	78/75		81/82
mol % G + C	45	39	35	38	38		38	35		37
Anabaena 90 mol % G + C	41	38	37	40	38	34	39	36	38	39

[a]References for the sequences: *Microcystis aeruginosa* PCC7806, Tillett et al., 2000; *M. aeruginosa* K-139, Nishizawa, et al., 2000; *M. aeruginosa* UV027, Raps et al., unpublished, AF458094; *Planktothrix agardhii* CYA126/8, Christiansen et al., 2003.

[0174] When the microcystin synthetase genes were compared to the anabaenopeptilide synthetase genes of *Anabaena* 90, the highest similarity, 54%, was between mcyC and apdD.

[0175] In the genome databases of *Anabaena* 7120 (http://www.kazusa.or.jp/cyano/*Anabaena*/search.html) and *Nostoc punctiforme* (http://www.igi.doe.gov/JGI_microbial/html/nostoc/nostoc_homeoage.html) no genes were found with more than 50% identity to the microcystin synthetase genes at the amino acid level. There are two sequences in the genome database of *Anabaena/Nostoc* 7120 named "microcystin synthetase B" on account of similarity to mcyB of *Microcystis aeruginosa* (AY034602): all2643 (ID:3312, 3309 bp) and a112647 (ID:3317, 3261 bp), (identity: 47.0%, positive: 65.5% and identity: 43.9%, positive: 61.9%, respectively). The matches of these sequences with mcyB of *Anabaena* 90 are 53% and 51% at the gene level. The translated peptides are 49%/66% and 43%/61% identical/similar, respectively.

[0176] The G+C content of the microcystin synthetase gene cluster (56 kb) from *Anabaena* 90 is 39%, is lower than the value, 43%, for the region of the anabaenopeptilide synthetase (39 kb) (Rouhiainen et al., 2000). These figures are in the limits of the mol % G+C values 43.9, 39.1 and 42.3 for the type strains *Anabaena cylindrica* (PCC 7122), *Anabaena flos-aquae* (1?CC 9332) and for the reference strain of *Anabaena* cluster 2 (PCC 7108), respectively (Rippka et al., 2001).

Substrate Specificity of the Adenylation Domains

[0177] The substrate specificity-conferring amino acids in the adenylation domains of the microcystin synthetases of *Anabaena* 90, *P. agardhii* CYA126, *M. aerginosa* PCC7806, K-139, and UV027 were assessed according to Stachelhaus et al., (1999) (Table 2). The substrate specificity codes of the modules McyA-1, McyA-2, McyB-2 and of the nonribosomal peptide synthetase (NRPS) modules in McyG and mcyE are identical or nearly identical in all the sequenced microcystin synthetases (able 2).

TABLE 2

Specificity-conferring amino acids (signature sequences) of the adenylation domains in the microcystin synthetases from different cyanobacterial strains.

Module	Strain	Signature sequence[a]	Precedent SS	Activated amino acid	Reference template
McyA	*Anabaena* 90	**DVWHISLID**	**DVWHLSLID**	Ser	SyrE (1, 2)[b]
	M. aeruginosa 7806	**DVWHFSLID**	**DVWHFSLVD**		EntF, MycC (1, 23)[b]
	M. aeruginosa K-139	**DVWHFSLID**			
	M. aeruginosa UV027	**DVWHFSLID**			
	P. agardhii CYA 126/8	**DVWHISLID**			
McyA 2	*Anabaena* 90	**DLFNNALTY**			
	M. aeruginosa 7806	**DLFNNALTY**			
	M. aeruginosa K-139	**DLFNNALTY**	**DLFNNALTY**	Ala	BlmIX, MxA (4, 5)[c]
	M. aeruginosa UV027	**DLFNNALTY**			
	P. agardhii CYA 126/8	**DLFNNALSY**			
McyB 1	*Anabaena* 90	**DVWFFGLVD**			
	M. aeruginosa 7806	**DAWFLGNVV**	**DAWFLGNVV**	Leu	BacA, LicA, LicB,
	M. aeruginosa K-139	**DAWFLGNVV**			SrFA (1)[b]
	M. aeruginosa UV027	**DVWTIGAVE**		(Arg)	
	P. agardhii CYA 126/8	**DALFFGLVD**			
MCyB 2	*Anabaena* 90	**DARHVGIFV**			
	M. aeruginosa 7806	**DARHVGIFV**			
	M. aeruginosa K-139	**DABHVGIFV**	no precedents	(Asp/MeAsp)	

TABLE 2-continued

Specificity-conferring amino acids (signature sequences) of the adenylation domains in the microcystin synthetases from different cyanobacterial strains.

Module	Strain	Signature sequence[a]	Precedent SS	Activated amino acid	Reference template
	M. aeruginosa UV027	**DARHVGIFV**			
	P. agardhii CYA 126/8	**DPRHVGIF**I			
McyC	*Anabaena* 90	**DVWCFGLVD**			
	M. aeruginosa 7806	**DVW**TI**GAVD**			
	M. aeruginosa K-139	**DVW**TI**GAV**E	no precedents	(Arg)	
	M. aeruginosa UV027	**DVW**TI**GAVD**			
	P. agardhii CYA 126/8	**D**P**W**G**FGLVD**			
McyG	*Anabaena* 90	**GAFWVAASG**			
	M. aeruginosa 7806	**GAFWVAASG**	no precedents		
	M. aeruginosa K-139	**GAFWVAASG**			
	P. agardhii CYA 126/8	**GAFWVAASG**			
McyE	*Anabaena* 90	**DPRHSGVVG**			
	M. aeruginosa 7806	**DPRHSGVVG**	no precedents	(Glu)	
	M. aeruginosa K-139	**DPRHSGVVG**			
	P. agardhii CYA 126/8	**DPRHSGVVG**			

[a]Nine variable amino acids of the signature sequences determined as described by Stachelhaus et al., 1999. Bold letters indicate the residues, which are identical with the amino acids of the signature sequence from *Anabaena* 90.

[b]1. Stachelhaus et al., 1999, 2. Challis et al., 2000, 3. Duitman et al., 1999.

[c]4. Du et al., 2000, 5. Silakowski et al., 2001.

[0178] There are, however, more differences in the specificity codes of variable amino acids activating McyB-1 and McyC module. The substrate specificity regions of the adenylation domains (corresponding amino acids 235-331 of GrsA, Stachelhaus et al., 1999) in McyA, McyB and in McyC from *Anabaena* 90, *P. agardhii* and from M aeruginosa were compared by using the algorithm of Smith and Waterman in the EMBOSS program package. The substrate specificity regions of McyA, of the second module of McyB (McyB-2) and of McyC are highly conserved. In *Anabaena* 90 and *M. aeruginosa*, the identity/similarity values are 80/90% for McyA, 86/92% for McyB-2 and 70/80% for McyC. Between *Anabaena* 90 and *P. agardhii* the identity/similarity for the substrate specificity region of McyC is higher, 85/88%, but lower for the second module of McyA, 73/83%. The substrate specificity region of McyB-1 is considerably less conserved between *Anabaena* 90 and *M. aeruginosa* PCC7806, 29/53% than between *Anabaena* 90 and M aeruginosa UV027, or *P. agardhii,* 66/80%.

Activities Encoded by mcyG, mcyD and mcyE of *Anabaena* 90

[0179] Motif scan at Prosite (Database of protein families and domains) and at Pfam (Protein families) database (http://hits.isb-sib.ch/cgi-bin/PFSCAN) and Conserved Domain (CD) search at NCBI (http://www.ncbi.nlm.nih.gov/Structure/cdd/wrpsb.cgi) were used to discover the putative functions of McyG, mcyD and McyE. In the N-terminal part of McyG a NRPS module was identified, which contains an adenylation domain and a thiolation (phosphopantetheine carrier) domain. Next to this, toward the C-termiinus there are four polyketide synthase (PKS) domains: β-ketoacyl synthase (KS), acyltransferase (AT), ketoreductase (KR) and acylcarrier protein (ACP), in this order. Between AT and KR domains there is a C-methyltransferase, MeT/CM) domain (FIG. **1**). McyD contains two modules of the type I polyketide synthases. The first module consists of KS, AT, dehydratase (DH), MeT (CM) (FIG. **1**), KR and ACP domains; and module two has KS, AT, DK KR and ACP domains, in the presented orders. mcyE is the other mixed PKS/NRPS, including PKS domains KS, AT, ACP and MeT (CM) (FIG. **1**; FIG. **2**A). These are followed by a unique aminotransferase domain (AMT) (FIG. **1**; FIG. **2**B) found in other microcystin synthetases (Tillet et al., 2000; Christiansen et al., 2003), and also in the synthetases of mycosubtilin (Duitman et al., 1999) and iturin (Tsuge et al., 2001) of *Bacillus subtilis*. At the N-terminal region, subsequently there is a NRPS module comprising of two condensation domains, an adenylation and a thiolation (peptidyl carrier) domain (FIG. **1**).

Ketoreductase and Dehydratase Domains

[0180] The activity of the KR domains of McyG (one) and McyD (two) can be predicted from the microcystin synthetases structure, and they have the NAD cofactor binding motif, $GXGXX(G/A)(X)_3(G/A)M(X)_6G$, common to oxidoreductases (Scrutton et al., 1990). (FIG. **3**B) The DH domains in the modules of McyD (AMCD-DH2 and AMCD-DH3) contain the active site motif $H(X)_3D(X)_4P$ and $H(X)_3G(X)_4P$, respectively (FIG. **3**A). The motif in AMCD-DH3 is identical to the consensus sequence (Aparicio et al., 1996). The motif $H(X)_3D(X)_4P$, where Gly is substituted by Asp, is also found in the active DH domain of module 10 in rifamycin synthase (Tang et al., 1998) (FIG. **3**). This supports the conclusion based on the microcystin structure, that the DH domains in McyD are functional.

Specificity of the Acyl Transferase Domains

[0181] From the structure of the microcystins it is possible to conclude that the single AT domains of McyG and McyE, and the first AT domain of McyD, load methylmalonyl-CoA. But the presence of methyltransferase domains in McyG, McyD and mcyE wig. 1, FIG. **2**A) suggests that the loading unit can be malonyl-CoA. Regions have been identified in AT domains, where the sequences are different depending on the specificity for either malonyl-CoA or methylmalonyl-CoA (FIG. **4**) (Ikeda et al., 1999). By analysing the

sequences of the acyltransferase domains (FIG. **4**) and comparing them with the AT domains of soraphen and rapamycin synthases, which utilize malonyl subunits, we concluded that all the AT domains of microcystin synthetase load malonyl units. The methyltranferase domains of McyG, McyD and mcyE carry out three methylations in the positions indicated with arrows (FIG. **1**). The CD search relates these domains to the UbiE/COQ5 C-methyltranferase family.

Ketosynthase and Acylcarrier Protein Domains

[0182] The active site cysteine and the two histidine residues which are present in polyketide synthases (Aparicio et al., 1996) were found in the KS domains of McyG, McyD and McyE (FIG. **5**A). The only ACP domain of McyG and the first ACP domain of McyD have the active site sequence MGXDS, where a methionine residue replaces the commonly identified leucine residue (FIG. **5**B). There are also variations in this position of the rifamycin synthase (rang et al., 1998). The ACP domain from the second module of McyD has the active site motif LGLNS (FIG. **5**B), where Asn takes the place of the generally found Asp as in the module 11 of the rapamycin synthase (Aparicio et al., 1996).

The Order of the Genes in the Microcystin Synthetase Gene Cluster is Different in the Cyanobacterial Species

[0183] The arrangement of the genes is different in the gene clusters of microcystin biosynthesis from the strains of three species. In *Anabaena* strai 90, *Microcystis* aeruginosa (Tillett et al., 2000; Nishizawa et al., 2000) and in *Planktothrix agardhii* CYA126 (Christiansen et al., 2003) the NRPS genes, mcyA, mcyB and mcyC have the same order, but the organization of the other genes is different. In *Anabaena* strain 90 and in *M. aeruginosa* the mcy-genes are in two clusters, which are transcribed in opposite directions, whereas in *P. agardhii* they are in one cluster transcribed in the same direction (except mcyT, which was not found in *Anabaena* and *Microcystis*). The arrangement of the genes from mcyD to mcyH in *Microcystis* is almost identical in *Planktothrix* (mcyF is missing in *Planktothrix*), but it differs from the order in *Anabaena*. In *Planktothrix*, compared to *Microcystis*, the part containing mcyD, mcyE, mcyF, mcyG, mcyH, mcyI and mcyJ is reversed. In this rearrangement, mcyF and mcyI were lost from the cluster and mcyJ was relocated after mcyC.

The Biosynthesis of Microcystins

[0184] In *Anabaena*, the order of the domains coded by the genes in the two sets is co-linear with the hypothetical sequence of the enzymatic reactions for microcystin biosynthesis (FIG. **1**). The progression of the biosynthetic reactions follows the order of the functions coded first by mcyG and continuing with the activities coded by mcyD, mcyJ, mcyE, mcyF, mcyI, mcyA, mcyB and mcyC.

[0185] Phenyl acetate is the assumed staring unit in the biosynthesis of Adda (Moore et al., 1991). It is activated by the adenylating domain identified in the N-terminus of McyG, and transferred onto the subsequent thiolation (phosphopantetheine binding) site. Polyketide synthesis reactions are followed (FIG. **1**). All four extension units are malonyl-CoA molecules according to the substrate specificity of the AT domains (FIG. **4**). In McyG there is a KS domain to catalyse the first condensation reaction between phenylacetate and malonyl-CoA.

[0186] The reductive reactions needed to fashion the polyketide chain are putatively catalysed by KR and DH domains of McyD and McyE. The KR domain of McyG is in the right position to reduce the carbonyl group of the putative starter molecule. The methyltransferase domains of McyG, McyD and mcyE are the obvious candidates to introduce three methyl groups into the carbon frame of Adda. It was recently verified with a knockout mutant (Christiansen et al., 2003) that the incorporation of the fourth methyl, which is seen in the methoxy group of Adda, is catalysed by McyJ. The amino transferase domain of mcyE most likely adds the amino group, which participates in the peptide bond with the glutamate residue.

[0187] There are two condensation domains of peptide synthetases in McyE. The first one logically catalyses the peptide bond between Adda and glutamate, which is activated by the adenylation domain of McyE. The signature sequence, which was also determined as DPRHSGVVG for McyE of both *M. aeruginosa* and *P. agardhii*, has no precedents in the databases (Table 2). The synthetases of other peptides, which contain glutaryl residues are known for bacitracin, fengycin and surfactin (accession numbers: AF007865, AF023464, AF087452 and D13262). In these compounds the standard α-carboxyl of glutamate is part of the peptide bond, while in microcystins it is the γ-carboxyl. This is analogous to the activation of aspartate/methylaspartate by the second adenylation domain of McyB. The β-carboxyl of aspartate/methylaspartate instead of the α-carboxyl is engaged in the peptide bond formation. This must have impact on the compositions of the glutamate and aspartate/methylaspartate binding pockets in the adenylation domains.

[0188] McyA has two adenylation domains for the activation of serine and alanine, respectively. The signature sequences of these domains have models and are almost identical in *Anabaena* 90, *M. aeruginosa* and *P. agardhhi* (Table 2). The dehydration of serine supposedly takes place after the activation by adenylation and is catalysed by McyI, which is similar to phosphoglycerate dehydrogenases.

[0189] There is only one, internal, condensation domain in McyA, which most likely links dehydroserine and D-alanine. The bond between glutamate and dehydroserine is putatively catalysed by the C-terminal condensation domain of McyE. There is a methyltransferase domain in the first module of McyA for N-methylation of dehydroserine. The epimerase domain at the C-terminus of McyA converts L-alanine to the D-form.

[0190] Two modules of McyB and one module of McyC logically activate, and add three residues to the nascent peptide chain: L-leucine or L-arginine, methylaspartate or aspartate and L-arginine, respectively (FIG. **1**). The amino

acids activated by the adenylation domains of McyC and by the first module of McyB (McyB-1) vary most frequently in microcystins. *M. aeruginosa* PCC7806 and *M. aeruginosa* K-139 produce mainly Mcyst-LR, and the substrate specificity conferring sequences in McyB-1 of these strains are identical with the signature sequence for leucine (Table 2). *M. aeruginosa* UV027 and *P. agardhii* CYA126 produce mostly Mcyst-RR, which is also produced by *Anabaena* 90 together with Mcyst-LR. Their signature sequences in McyB-1 are different and have no precedents in the databases (Table 2). In *M. aeruginosa* UV027 the specificity codes of McyB-1 and McyC are almost identical (DVWTIGAVE/DWTIGAVD) and match with the codes of McyC from *M. aeruginosa* K-139 and *M. aeruginosa* PCC7806, respectively (Table 2). Accordingly McyB-1 of *M. aeruginosa* UV027 and McyC activate arginine.

[0191] There is no epimerase domain in McyB of *Anabaena* 90 or in the other sequenced versions of McyB, though in microcystins, the aspartyl or methylaspartyl moiety is in the D-form. The epimerization in this position and in the glutamyl residue is putatively catalysed by McyF, which in a BLAST search was similar to aspartate racemases, and was shown by Nishizawa et al., (2001) to complement a D-glutamate deficient mutant of *Eschericia coli*. The C-terminal thiosterase domain of McC, as generally in nonribosomal peptide synthesis, (Kohli et al., 2001) catalyzes the final step in microcystin biosynthesis, the cyclization of the linear peptide (FIG. **1**).

[0192] McyH is probably not needed for the synthesis of microcystins but it may participate in the transport of microcystins.

Example 2

Taxon Sampling, Amplification and Sequencing

[0193] Genomic DNA from 36 strains of *Anabaena, Microcystis, Planktothrix, Nodularia*, and *Nostoc* was extracted. We chose three regions of the microcystin synthetase gene cluster to study the evolution of this biosynthetic system in cyanobacteria. A fragment of 291-297 bp from the mcyA gene was amplified with mcyA-Cd 1R (5'-aaaagtgttttattagcggctcat-3') and mcyA-Cd 1F (5'-aaaattaaaagccgtatcaaa-3') primers and sequenced as described earlier (Hisbergues et al. 2003). An 818 bp region of the mcyD gene was amplified with mcyDF (5'-gatccgattgaattagaaag-3') and mcyDR (5'-gtattccccaagattgcc-3') primers. An 809-812 bp region of the mcyE gene was amplified with the mcyE-F2 (5'-gaaatttgtgtagaaggtgc-3') and mcyE-R4 (5'-aattctaaagcccaaagacg-3') primers. The mcyE PCR products of *Nodularia* sp. strains were cloned with the TOPO TA cloning kit (Invitrogen) according to the manufacturer's instructions. The rpoC1 gene fragment of 750 bp was amplified with degenerate primers RF (5'-tgggghgaaagnacaytncctaa-3') and RR (5'-gcaaancgtccnccatcyaaytgba-3'). PCR reactions for mcyE, mcyD and rpoC1 were performed in a 20 μl final volume containing 1 μl of DNA, 1×DynaZyme II PCR buffer, 250 μM of each deoxynucleotide, 0.5 μM of both PCR primers, and 0.5 U of DynaZyme II DNA polymerase (Finnzymes, Espoo, Finland). The following protocol was used: 95° C., 3 min; 30×(94° C., 30 sec; 56° C., 30 sec; 72° C., 1 min); 72° C., 10 min. A region containing the 16S rRNA gene and the internal transcribed spacer 1 (ITS1) was amplified using primers and conditions described earlier (Lepére et al., 2000) from strains, for which the 16S rRNA sequence data was not available. The mcyD and mcyE gene products were sequenced directly with primers used for amplification except for the cloned mcyE sequences of *Nodulria* sp. strains, which were sequenced with primers anchored in the pCR2.1-TOPO vector, M13F (−20) and M13R. The rpoC1 gene products were sequenced with the amplification primers and with two additional internal sequencing primers RintF (5'-gatatgcccctgcgggatgt-3') and RintR (5'-acatcccgcaggggcatatc-3'). The 16S rRNA gene region of the amplified PCR products was sequenced directly using sets of internal primers (Edwards et al., 1989).

[0194] Sequencing of the mcyD, mcyE and 16S rRNA genes was performed by Genome Express (France). The rpoC1 products were sequenced with ABI PRISM 310 Genetic Analyzer. The mcyA sequences were assembled as described by Hisbergues at al. The chromatograms of mcyD, mcyE, rpoC1 and 16S rRNA gene sequences were checked and edited with Chromas 2.2 program (Technelysium Pty Ltd.). Contig assembly and alignment of the sequences were performed with BioEdit Sequence Alignment Editor (Hall et al., 1999).

Phylogenetic Analyses

[0195] Primer sequences and ambiguous regions of the alignments were excluded. The aligned data sets were the following lengths: mcyA (99 amino acids), mcyD (286 amino acids), mcyE (270 amino acids), rpoC1 (750 bp) and 16S rRNA (1455 bp). These sequences were combined with the sequence available from *Microcystis aeruginosa* PCC 7806 (Tillett et al., 2000). and *Planktothrix agardhii* NIVA-CYA 126/8 (Christiansen et al., 2003).

[0196] Outgroups for each of the three microcystin synthetase genes were identified with BLAST searches (Supplementary Information). We aligned mcyA, mcyE, and mcyD and the top three hits in BLAST searches with BioEdit (Hall et al. 1999).

[0197] Only conserved and reliably aligned sequence regions from the outgroup sequences were used in order to minimise potential phylogenetic reconstruction artefacts derived from the use of distant outgroups Swofford et al. 1996). In order to assess the stability of the ingroup tree topology, which could be influenced by the addition of outgroup lineages due to long branch attraction, the phylogenetic trees were analysed with and without the chosen outgroups. Phylogenetic analyses were performed with PAUP (Swofford, 2001) and PHYLIP (Felsenstein, 1993). Maximum likelihood and maximum parsimony analyses were used to reconstruct trees from each mcy gene fragment, and to compare the tree topologies of the separate and concatenated mcy gene sets and the 16S rRNA and rpoC1 genes. 16S rRNA sequences of 53 cyanobacterial strains and three outgroup species were used to construct a maximum-likelihood tree, to which the distribution of microcystin and nodularin producing cyanobacteria among other cyanobacteria was mapped (FIG. **8**).

TABLE 3

Accession numbers for sequences used in phylogenetic reconstruction. A solid line denotes unsuccessful attempts to amplify this region from the three strains of the genus *Nodularia* used in this study. A dash indicates cases where no attempt was made to obtain sequence data.

Taxon	mcyA	mcyE	mcyD	16S rRNA	rpoC1
Microcystis sp. HUB 5-2-4	AJ515451	—	—	—	—
Microcystis aeruginosa NIES 89	AJ515459	AY382530	AY424988	U03403	—
Microcystis sp. 199	AJ515452	—	—	AJ133172	—
Microcystis sp. GL260735	AJ515454	AY382531	—	AY439282	—
Microcystis sp. GL280646	AJ515455	AY382532	—	—	—
Microcystis sp. IZANCYA5	AJ515456	AY382533	—	—	—
Microcystis sp. IZANCYA25	—	AY382534	—	—	—
Microcystis sp. TuM7C	AJ515458	—	—	—	—
Microcystis viridis NIES 102	AJ515457	AY382535	AY424991	U40332	AY425001
Microcystis aeruginosa PCC 7941	AJ515460	AY382536	AY424989	U40340	—
Microcystis aeruginosa PCC 7806	AF183408	AF183408	AF183408	AF139299	AY425000
Microcystis sp. 98	—	AY382537	—	—	—
Microcystis sp. 205	AJ515453	AY382538	AY424990	AY439281	—
Nostoc sp. 152	AJ515475	AY382539	AY424984	AJ133161	AY424997
Nodularia spumigena HEM	_____	AY382540	AY424985	AF268005	AY424999
Nodularia spumigena BY1	_____	AY382541	AY424987	AF268004	—
Nodularia sp. F81	_____	AY382542	AY424986	AY439283	AY424998
Anabaena sp. 66A	AJ515462	AY382543	AY424983	AJ133157	—
Anabaena sp. 66B	AJ515463	—	—	—	—
Anabaena flos-aquae NIVA-CYA83/1	AJ515466	AY382544	—	AJ133158	—
Anabaena sp. 202A1/35	AJ515464	AY382545	AY424980	AJ133159	—
Anabaena lemmermannii 202A2	AJ515465	AY382546	AY424981	AJ293104	AY424995
Anabaena sp. 90	AJ515461	AJ536156	AJ536156	AJ133156	AY424996
Anabaena sp. PH256	—	AY382547	—	—	—
Anabaena sp. 315	—	AY382548	—	—	—
Anabaena sp. 318	—	AY382549	—	—	—
Anabaena sp. 299	—	AY382550	AY424982	AJ293106	—
Planktothrix sp. HUB 076	AJ515472	—	—	—	—
Planktothrix sp. PCC7821	AJ515473	—	—	—	—
Planktothrix sp. NIVA-CYA34	AJ515474	—	—	—	—
Planktothrix sp. 49	AJ515470	AY382551	AY424992	AJ133167	AY425003
Planktothrix sp. 97	AJ515471	AY382552	—	—	—
Planktothrix sp. NIVA-CYA126	AJ441056	AJ441056	AJ441056	AJ133166	—
Planktothrix sp. NIVA-CYA127	AJ515468	AY382553	AY424993	AJ133168	AY425002
Planktothrix sp. NIVA-CYA128/R	AJ515469	AY382554	AY424994	AJ133169	—
Oscillatoria sp. 213	—	AY382555	—	—	—
Oscillatoria sp. 226	—	AY382556	—	—	—

[0198]

TABLE 4

Accession numbers of sequences used to root the microcystin gene data set in FIG. 7. The outgroup sequences identified by BLAST searches were fused together to form three outgroup sequences in the mcyA, mcyD, and mcyE concatenated gene data set.

Gene	Outgroup	Accession	Organism	Gene	Function
McyA	Outgroup 1	AF210249	*Streptomyces verticillus*	blmX	Bleomycin biosynthetic gene
	Outgroup 2	AE004755	*Pseudomonas aeruginosa*	PA3327	Probable non-ribosomal peptide synthetase
	Outgroup 3	X97860	*Amycolatopsis mediterranei*	aps	Peptide-synthetase
McyD	Outgroup 1	AF395828	*Aphanizomenon ovalisporum*	aoaC	Polyketide synthase
	Outgroup 2	AJ421825	*Stigmatella aurantiaca*	stiH	Stigmatellin biosynthetic gene
	Outgroup 3	AP003590	*Nostoc* sp. PCC 7120	alr2680	Polyketide synthetase
McyE	Outgroup 1	D29676	*Bacillus brevis*	Grs2	Gramicidin S synthetase 2
	Outgroup 2	X70356	*Bacillus subtilis*	sifA1	Surfactin synthetase
	Outgroup 3	AF004835	*Brevibacillus brevis*	tycC	tyrocidine synthetase 3

[0199]

TABLE 5

Accession numbers for 16S rRNA sequences used to construct the maximum-likelihood tree presented in FIG. 8.

Species	Strain	16S rRNA
Cyanobacteria		
Subsection I *Chroococcales*		
Cyanobium gracile	PCC 6307	AF001477
Cyanothece sp.	PCC 7424	AF132932
Gloeobacter violaceus	PCC 7421	AF132790
Gloeothece membranacea	PCC 6501	X78680
Microcystis aeruginosa	PCC 7806	U03402
Microcystis aeruginosa	PCC 7941	U40340
Microcystis wesenbergii	NIES 104	AJ133174
Synechococcus elongatus	PCC 6301	X03538
Synechococcus leopoliensis	PCC 7942	AF132930
Synechococcus sp.	PCC 7002	AJ000716
Synechococcus sp.	PCC 6716	AF216942
Synechococcus sp.	WH 8103	AF311293
Synechocystis sp.	PCC 6803	D64000
Thermosynechococcus elongatus	BP-1	AP005376
Prochlorococcus marinus	MED 4	AF001466
Prochlorococcus marinus	MIT 9313	AF053399
Subsection II *Pleurocapsales*		
Chroococcidiopsis sp.	SAG 2023	AJ344552
Chroococcidiopsis thermalis	PCC 7203	AB039005
Myxosarcina sp.	PCC 7312	AJ344561
Myxosarcina sp.	PCC 7325	AJ344562
Pleurocapsa minor	SAG 4.99	AJ344564
Pleurocapsa sp.	PCC 7516	X78681
Xenococcus sp.	PCC 7305	AF132783
Subsection III *Oscillatoriales*		
Arthrospira sp.	PCC 8005	X70769
Leptolyngbya sp.	PCC 7375	AF132786
Leptolyngbya sp.	PCC 7104	AB039012
Limnothrix redekei	NIVA-CYA 227/1	AB045929
Lygnbya aestuarii	PCC 7419	AJ000714
Oscillatoria rosea	IAM-220	AB003164
Oscillatoria sancta	PCC 7515	AF132933
Planktothrix agardhii	NIVA-CYA 126	AJ133166
Planktothrix sp.	2	AJ133185
Planktothrix sp.	49	AJ133167
Pseudanabaena sp.	PCC 6903	AF132778
Spirulina major	PCC 6313	X75045
Spirulina subsalsa	IAM-223	AB003166
Trichodesmium erythraeum	IMS101	Unpublished*
Prochlorothrix hollandica	—	AF132792
Subsection IV *Nostocales*		
Anabaena sp.	66A	AJ133157
Anabaena sp.	90	AJ133156
Anabaenopsis circularis	NIES 21	AF247595
Anabaenopsis sp.	PCC 9215	AY038033
Aphanizomenon flos-aquae	NIES 81	AJ293131
Cyanospira rippkae	PCC 9501	AY038036
Cylindrospermum stagnale	PCC 7417	AF132789
Nodularia spumigena	BY1	AF268004
Nodularia sp.	F81	AY439283
Nodularla spumigena	PCC 73104	AF268023
Nostoc sp.	PCC 7120	X59559
Nostoc punctiforme	PCC 73102	AF027655
Nostoc sp.	152	AJ133161
Nostoc sp.	PCC 9709	AF027654
Scytonema hofmannii	PCC 7110	AF132781
Subsection V *Stigonematales*		
Chlorogloeopsis sp.	PCC 7518	X68780
Fischerella muscicola	PCC 7414	AP132788

TABLE 5-continued

Accession numbers for 16S rRNA sequences used to construct the maximum-likelihood tree presented in FIG. 8.

Species	Strain	16S rRNA
Outgroups		
Bacillus subtilis	BS62	AB016721
Chlorobium tepidum	—	M58468
Escherichia coli	K12	AE000129

*Unpublished 16S rRNA obtained from *Trichodesmium erythraeum* IMS101 on the Joint Genome Institute webpage (www.jgi.doe.gov).

Example 3

[0200] Primer design and specificity testing. General microcystin synthetase E forward primer (mcyE-F2) and genus specific reverse primers for *Anabaena* (AnamcyE-12R) as well as for Microcystis (MicmcyE-R8) (able 6) were designed with mcy gene sequences of *Anabaena* 90 (see Example 1), by using BLAST (1) and BioEdit (Hall 1999).

[0201] Specificity of these primers was tested with 14 *Anabaena,* 13 *Microcystis,* 8 *Planktothrix* strains and with one *Nostoc* strain (Table 7). *Microcystis* and *Planktothrix* strains were grown in Z8 medium (Kotai 1972), whereas *Anabaena* and Nostoc strains were grown in a modified Z8 medium without nitrogen. The strains were grown under continuous light (20 μmol $m^{-2}s^{-1}$) at 20±2° C.

[0202] PCR reaction was carried out with 1 μl of extracted DNA, 1×DynaZyme II PCR buffer [10 mM Tris-HCl, pH 8.8 at 25° C., 1.5 mM MgC_2, 50 mM KCl, 0.1% Triton X-100, (Finnzymes)], 250 μM dNTPs (Finnzymes), 0.5 μM of primers (Sigma-Genosys Ltd.) and 0.5 U of DyNAzyme II DNA polymerase (Finnzymes) in a volume of 20 μL. The PCR amplification was performed with initial denaturation at 95° C. for 3 min followed by either 30 (*Anabaena*) or 25 (*Microcystis*) cycles at 94° C. for 30 s, at 58° C. for *Anabaena* and at 60° C. for *Microcystis* for 30 s and at 72° C. for 60 s, followed by 10 min final extension at 72° C. Presence or absence of the mcyE product was determined using 20 μl of amplification product and 1.5% agarose gel electrophoresis.

[0203] Lake water samples. Water samples were collected at Lake Tuusulanjärvi from 0 to 2 m depth every second or third week during summer period 1999. For DNA extraction one liter of lake water was concentrated to less than 2 ml by centrifugation and stored at −70° C. Lake Hiidenvesi consists of several natural basins representing a transition from hypertrophy to mesotrophy. Water samples were collected from 3 to 5 different depths from basins of Kirkkojärvi (3.5 m deep at the sampling site), Mustionselkä (4 m), Nummelanselkä (6 m), and Kiihkelyksenselkä (30 m) on 15 Aug. 2001. For DNA extraction 100 ml of lake water was filtered through 3 μm pore size Poretics® polycarbonate disc filter (47 mm), (Osmonics Inc.) and cells were stored with lysis buffer at −20° C. (14). For microcystin concentration analysis, 5 ml of lake water was stored in a glass vial at −20° C. Cyanobacterial cell densities were determined using the inverted microscope technique (Utermöhl, 1958) from the samples which were preserved with acid Lugol's solution (Willen, 1962) and stored in darkness at 4° C.

[0204] Isolation and purification of DNAs. Genomic DNAs of the *Anabaena, Microcystis, Planktothrix* and *Nos-*

toc strains and the lake water samples were extracted with a hot phenol-chloroform-isoamylalcohol-method (Giovannoni et al., 1990). Extracted DNAs were purified either once (strains) or twice (lake water samples) with Prep-A-Gene® DNA Purification Systems (Bio-Rad) according to the manufacturer's instructions and eluted in 60 μl.

[0205] QRT-PCR. External standards for mcyE copy number quantification were prepared using genomic DNAs of strains *Anabaena* 90, 315, and 202A1 as well as those of *Microcystis* GL 260735, PCC 7806, and PCC 7941. Genomic DNA concentration of these DNAs was measured with a spectrophotometer at 260 nm (Beckman DU-7400). Purity was determined by calculating the ratio of the absorbances measured at 260 nm and 280 nm. Approximate genome sizes, *Anabaena* 5.15 Mb and *Microcystis* 4.70 Mb, were used in mcyE copy number calculation. These genome sizes were estimated based on the genome sizes of *Anabaena* PCC 6309, *Anabaena* PCC 7122 and *Microcystis* PCC 7941 (Castenholz, 2001). The mcyE copy numbers of the standard strains DNAs were calculated using following equation with the assumption that each genome had only one mcyE gene and the molecular weight of one bp was 660 g mol^{-1}:

$$\text{Copies } \mu l^{-1} = \frac{6\times10^{23}[\text{copies mol}^{-1}]\times \text{DNA concentration } [g\ \mu l^{-1}]}{\text{Molecular weight of one genome } [g\ \text{mol}^{-1}]}. \quad \text{Equation 1}$$

[0206] Ten-fold dilution series of genomic DNAs of the standard strains were prepared and these dilutions were amplified with *Anabaena* and *Microcystis* mcyE QRT-PCR. Linear regression equations of the obtained cycle threshold values (Ct values, i.e. the first turning points of the fluorescence curves as a function of cycle numbers) were calculated as a function of known mcyE copy numbers.

[0207] The QRT-PCR reaction was carried out with 1 μl of DNA of standard strains or lake water samples, 3 mM $MgCl_2$, 0.5 μM of both primers (Sigma-Genosys Ltd.) and 1 μl of hot start reaction mix to a final volume of 10 μl (LightCycler—fastStart DNA master SYBR green I—kit, Roche Diagnostics). Amplification was performed with initial preheating of 10 min at 95° C. followed by 45 cycles at 95° C. for 2 s, at 58° C. for 5 s and at 72° C. for 10 s. Generation of the products was monitored after each extension step at 77° C. in *Anabaena* and 78° C. in *Microcystis* mcyE QRT-PCR by measuring fluorescence of double-stranded DNA binding SYBR green 1 dye using LightCycler QRT-PCR (Roche Diagnostics). All lake water samples were amplified three times. The Ct values were determined by the second derivative maximum method of LightCycler software (version 3.5). Copy numbers of mcyE gene of the lake water samples were determined by converting obtained Ct values into the mcyE copy numbers according to the regression equations of the external standards that gave the highest (*Anabaena* 202A1 and *Microcystis PCC*7941) and lowest (*Anabaena* 315 and *Microcystis* PCC7806) mcyE copy numbers (FIGS. **9**A and B).

[0208] Amplification efficiencies, e ($e=10^{-1/S}-1$, s=slope of the linear regression), of the *Anabaena* and *Microcystis* mcyE QRT-PCR with standard strains were calculated as a function of known mcyE copy numbers and with those of Lake Tuusulanjärvi DNA samples as a function of different dilutions of the samples.

[0209] In order to determine melting temperatures for the amplification products of the standard strains and of the lake water samples, temperature was raised after QRT-PCR from 65° C. to 95° C. and fluorescence was detected continuously. Characteristic melting temperatures of the mcyE QRT-PCR products were determined with LightCycler software (version 3.5).

[0210] Microcystin analysis of the strains and lake water samples. Dry weight of the *Anabaena, Microcystis, Planktothrix* and *Nostoc* strains was measured and microcystin was extracted by sonication as detailed previously (Repka et al., 2001). Microcystin concentration of the strains was analyzed with an Agilent 1100 Series high performance liquid chromatograph with a diode array detector and Luna 5 μm C18 column (150×2 mm, Phenomenex). A mobile phase was 10 mM ammonium acetate and acetonitrile. During 6 to 40 minutes, concentration of acetonitrile increased from 24% to 60%. Flow rate was 0.2 ml min^{-1} at 40° C., injection volume 20 μl, and detection at 238 nm. Purified microcystin-LR was used as a standard and microcystins were identified by their UV spectra and retention times.

[0211] Total microcystin of the lake water samples was extracted from 5 ml of lake water using tip sonicator for 5 min (Braun Labsonic-U). Prior measuring microcystin concentration with EnviroGard® microcystins plate kit (Strategic Diagnostics Inc.) and plate spectrophotometer (Labsystems iEMS reader MF) samples were filtered through 0.2 μm Puradisc™ filters (Whatman) to remove the particles.

[0212] Statistical analysis. Spearman correlation coefficients between microcystin concentration (μg l^{-1}), mcyE copy numbers (copies ml^{-1}), and *Anabaena* as well as *Microcystis* cell numbers (cells ml^{-1}) of lake water samples were calculated with SAS® statistical software for Windows (SAS Institute Inc.).

[0213] Specificity of the primers. The mcyE gene primers (Table 6) were both genus and mcyE gene specific, since a single amplification product was observed when genomic DNA of microcystin producing *Anabaena* or *Microcystis* strain was used as a template in PCR with *Anabaena* or *Microcystis* genus specific primers (Table 7).

[0214] Detection range of mcyE copy numbers. The QRT-PCR was log-linear from 6.6×10^2 to 6.6×10^5 mcyE copies in a reaction when the genomic DNAs of the standard strains *Anabaena* 90, *Anabaena* 202A1, *Microcystis* GL 260735 or *Microcystis* PCC 7941 were used as a template and from 6.6×10^3 to 6.6×10^6 when those of standard strains *Anabaena* 315 or *Microcystis* PCC 7806 were used (FIGS. **9**A and B). The lowest reliable mcyE copy numbers in Lake Tuusulanjärvi were 42, 84, 33, and 63 copies ml^{-1} when calculated with the regression equations of the standards *Anabaena* 315, *Anabaena* 202A1, *Microcystis* 7806, and *Microcystis* 7941. In Lake Hiidenvesi the lowest reliable mcyE copy numbers were ten times higher than in Lake Tuusulanjärvi, 420, 840, 330, and 630 copies ml^{-1} when calculated with the same standards, respectively. One ng of genomic DNA of *Anabaena* and *Microcystis* standard strains contained 1.76×10^5 and 1.94×10^5 mcyE copies. The purity of these DNAs varied from 1.8 to 1.9.

[0215] The mcyE copy numbers of lake water. *Microcystis* mcyE copy numbers in Lake Tuusulanjärvi were 11 to 91 times more abundant than those of *Anabaena* mcyE copy numbers calculated as a ratio of the average mcyE copy numbers obtained with *Anabaena* 315, *Anabaena* 202A1,

Microcystis PCC 7941 and *Microcystis* PCC 7806 standards (FIG. **10**). *Microcystis* mcyE copy numbers were also more abundant than those of *Anabaena* in the Basin of Kiihkelyksenselkä of Lake Hiidenvesi (FIG. **11**). In the Basins of Nummelanselkä and in MustionselkäMicrocystis and *Anabaena* mcyE copy numbers were quite similar (FIG. **11**). In the Basin of Kirkkojärvi both *Microcystis* and *Anabaena* mcyE copy numbers were below the detection limits determined with the standards (FIG. **11**). In Lake Hiidenvesi (FIG. **11**) the average mcyE copy numbers of *Anabaena* and *Microcystis* as well as microcystin concentrations were lower than in Lake Tuusulanjärvi (FIG. **11**). Microcystin concentration had a statistically significant positive correlation with *Microcystis* mcyE copy numbers of all studied samples within the mcyE copy number detection range determined with the standards (Table 8).

[0216] Amplification efficiency. With Lake Tuusulanjärvi water samples the *Microcystis* mcyE QRT-PCR amplification efficiencies (0.78-0.99, Table 4) were similar to the amplification efficiencies of the *Microcystis* standards (0.86-0.94, Table 4). However, *Anabaena* mcyE QRT-PCR amplification efficiencies with Lake Tuusulanjärvi water samples (1.14 to 2.36, Table 4) were unrealistic high compared to the amplification efficiencies of the *Anabaena* standard strains (0.96-0.99, Table 9).

[0217] Melting curve analysis. Characteristic melting temperatures of the mcyE QRT-PCR products (247 bp) of the three *Anabaena* (average=79.6° C., CV=0.4%, n=38, Table 5) and three *Microcystis* (average=81.5° C., CV=0.2%, n=38, Table 5) standard strains corresponded to the melting temperatures of *Anabaena* (average=79.3° C., CV=0.3%, n=58) and *Microcystis* (average=81.7° C., CV=0.2%, n=63) mcyE QRT-PCR products amplified with lake water samples (data not shown). The 1.9° C. difference in the average characteristic melting temperatures was due to over 40 nucleotide difference between *Anabaena* and *Microcystis* mcyE sequences.

[0218] Primer dimers were detected in *Anabaena* and in *Microcystis* mcyE QRT-PCR with negative controls and in *Anabaena* mcyE QRT-PCR with lake water samples that had low template DNA concentration, although hot start Taq DNA polymerase provided by the manufacturer of the kit was used. The error caused by the primer dimers was avoided by measuring fluorescence of *Anabaena* and *Microcystis* mcyE QRT-PCR amplification at higher temperature (77° C., 78° C., respectively) than the melting temperature of the primer dimers.

[0219] Microcystin concentration and cyanobacterial cell density of lake water. Microcystin concentrations as well as *Anabaena* and Microcyptis cell densities were highest in Lake Tuusulanjärvi on July and started to decrease thereafter (FIGS. **10** and **12**). In Lake Hiidenvesi microcystin concentrations and cell densities were lower than those in Lake Tuusulanjärvi (FIGS. **11** and **13**). According to microscope analysis, *Microcystis* cells were more abundant than *Anabaena* cells in Lake Tuusulanjärvi whereas *Microcystis* cells were observed only occasionally in Like Hiidenvesi. *Anabaena* was the most dominant genus in the Basins of Kirkkojärvi and Mustionselkä of Lake Hiidenvesi whereas *Aphanizomenon* was the most dominant genus in the Basins of Nummelanselkä and Kiihkelyksenselkä of Lake Hiidenvesi as well as in the Lake Tuusulanjärvi.

TABLE 6

Primers used in this study.

Primer	Sequence (5' to 3')	
mcyE-F2	GAA ATT TGT GTA GAA GGT GC *	(SEQ ID NO 64)
AnamcyE-12R	CAA TCT CGG TAT AGC GGC	(SEQ ID NO 65)
MicmcyE-R8	CAA TGG GAG CAT AAC GAG	(SEQ ID NO 66)

* Forward primer, mcyE-F2, used in this study, was described in Example 2

[0220]

TABLE 7

Specificity of *Anabaena* (mcyE-F2, AnamcyE-12R) and *Microcystis* (mcyE-F2, MicmcyE-R8) microcystin synthetase E (mcyE) primers was studied using *Anabaena, Microcystis, Planktothrix,* and *Nostoc* strains. Presence (+) or absence (−) of the mcyE product. Microcystin (MC) production (+) or lack of production (−). Accession numbers indicate mcyE sequences available in GenBank. Culture collections: PCC, Pasteur Culture Collection, Paris, France; NIVA-CYA, Norwegian Institute for Water Research, Oslo, Norway; NIES, National Institute for Environmental Studies, Tsukuba, Japan.

Genus Strain	MC	*Anabaena* mcyE primers	*Microcystis* mcyE primers	Accession No	Reference
Anabaena					
66A	+	+	−	XX	47, b
90	+	+	−	AJ536156	47, a
202A1	+	+	−	XX	47, b
202A2/41	+	+	−	XX	47, b
NIVA-CYA83/1	+	+	−	XX	47, b
315	+	+	−	XX	b
318	+	+	−	XX	b
86	−	−	−		46
123	−	−	−		46
14	−	−	−		46
PCC 6309	−	−	−		43
PCC 7108	−	−	−		43
PCC 73105	−	−	−		43
PCC 9208	−	−	−		43

TABLE 7-continued

Specificity of *Anabaena* (mcyE-F2, AnamcyE-12R) and *Microcystis* (mcyE-F2, MicmcyE-R8) microcystin synthetase E (mcyE) primers was studied using *Anabaena, Microcystis, Planktothrix,* and *Nostoc* strains. Presence (+) or absence (−) of the mcyE product. Microcystin (MC) production (+) or lack of production (−). Accession numbers indicate mcyE sequences available in GenBank. Culture collections: PCC, Pasteur Culture Collection, Paris, France; NIVA-CYA, Norwegian Institute for Water Research, Oslo, Norway; NIES, National Institute for Environmental Studies, Tsukuba, Japan.

Genus Strain	MC	*Anabaena* mcyE primers	*Microcystis* mcyE primers	Accession No	Reference
Microcystis					
98	+	−	+	XX	47, b
205	+	−	+	XX	47, b
GL 260735	+	−	+	XX	55, b
GL 280646	+	−	+	XX	55, b
IZANCYA5	+	−	+	XX	53, b
IZANCYA25	+	−		XX	53, b
NIES102	+	−		XX	29, b
NIES A89	+	−	+	XX	29, b
PCC 7941	+	−	+	XX	43, b
PCC 7806	+	−	+	AF183408	43, 51
130	−	−	−		44
269	−	−	−		44
GL 060916	−	−	−		55
Planktothrix					
49	+	−	−	XX	47, b
97	+	−	−	XX	47, b
213	+	−	−		47
NIVA-CYA 126	+	−	−	AJ441056	9, 47
NIYA-CYA 127	+	−	−	XX	47, b
NIVA-CYA 128/R	+	−	−	XX	47, b
45	−	−	−		44
PCC 6304	−	−	−		43
Nostoc					
152	+	−	−	XX	48, b

a Example 1
b Example 2

[0221] (9) Christiansen et al. 2003, (29) Lyra et al., 2001, (43) Rippka and Herdman, 1992, (44) Rouhiainen et al. 1995, (46) Sivonen and Jones, 1999, (47) Sivonen et al. 1989, (48) Sivonen et al. 1995, (53) Vasconcelos et al., 1995, (55) Vezie et al. 1998,

TABLE 8

Spearman correlation coefficients between microcystin concentration (μg l^{-1}) and microcystin synthetase E (mcyE) copy numbers (copies ml^{-1}) calculated using different standards (*Anabaena* 202A1, *Anabaena* 315, *Microcystis* PCC 7806 and *Microcystis* PCC7941) and cell numbers (cells ml^{-1}) in Lake Tuusulanjärvi and Lake Hiidenvesi. Sum of *Anabaena* and *Microcystis* mcyE copy numbers was counted by adding the average copy numbers calculated using the two *Anabaena* and *Microcystis* standards. Number inside the parenthesis shows the number of samples used to calculate the spearman correlation.

	Anabaena McyE		*Microcystis* mcyE		Sum of *Anabaena and Microcystis* mcyE	*Microcystis* cells	*Anabaena* cells	*Microcystis Anabaena* cells
Lake water samples	202 A1	315	PCC 7806	PCC 7941				
All samples			0.57*	0.57*	0.52, p= 0.10			
	(11)	(11)	(15)	(15)	(11)	(21)	(21)	(21)
Lake Tuusulanjärvi	1*** (5)	1*** (5)	(6)	(6)	(5)	(7)	(7)	0.86 * (7)

TABLE 8-continued

Spearman correlation coefficients between microcystin concentration (μg l^{-1}) and microcystin synthetase E (mcyE) copy numbers (copies ml^{-1}) calculated using different standards (*Anabaena* 202A1, *Anabaena* 315, *Microcystis* PCC 7806 and *Microcystis* PCC7941) and cell numbers (cells ml^{-1}) in Lake Tuusulanjärvi and Lake Hiidenvesi. Sum of *Anabaena* and *Microcystis* mcyE copy numbers was counted by adding the average copy numbers calculated using the two *Anabaena* and *Microcystis* standards. Number inside the parenthesis shows the number of samples used to calculate the spearman correlation.

	Anabaena McyE		*Microcystis* mcyE		Sum of *Anabaena and Microcystis* mcyE	*Microcystis* cells	*Anabaena* cells	*Microcystis Anabaena* cells
Lake Hiidenvesi	(6)	(6)	(9)	(9)	(6)	(14)	(14)	(14)

*p < 0.5,
**p < 0.1,
***p < 0.01

[0222]

TABLE 9

Anabaena and *Microcystis* mcyE QRT-PCR amplification efficiencies, e ($e = 10^{-1/S} - 1$, S = slope of linear regression equation), of the external standard strains calculated as a function of mcyE copy numbers and those of Lake Tuusulanjärvi water samples calculated as a function of different dilutions of the samples. r^2 denotes coefficient of determination.

Strain or Sampling date	Amplification efficiency	S	r^2	mcyE copy numbers or Dilution factors
Microcystis				
GL 260735	0.86	−3.71	1	6.6×10^2, 6.6×10^3, 6.6×10^4, 6.6×10^5
PCC 7806	0.92	−3.53	1	6.6×10^2, 6.6×10^3, 6.6×10^4, 6.6×10^5, 6.6×10^6
PCC 7941	0.94	−3.47	1	6.6×10^2, 6.6×10^3, 6.6×10^4, 6.6×10^5
12-Jul	0.95	−3.46	1	1, 0.1, 0.05, 0.01, 0.005
2-Aug	0.97	−3.39	1	1, 0.1
23-Aug	0.99	−3.34	1	1, 0.1
7-Sep	0.80	−3.92	1	1, 0.1
20-Sep	0.78	−3.99	1	1, 0.1
6-Oct	0.88	−3.66	1	1, 0.1
Anabaena				
90	0.96	−3.41	1	6.6×10^2, 6.6×10^3, 6.6×10^4, 6.6×10^5
315	0.99	−3.34	1	6.6×10^2, 6.6×10^3, 6.6×10^4, 6.6×10^5, 6.6×10^6
202A1	0.98	−3.36	1	6.6×10^2, 6.6×10^3, 6.6×10^4, 6.6×10^5
12-Jul	1.32	−2.74	1	1, 0.1, 0.05
2-Aug	1.14	−3.02	1	1, 0.1, 0.05
23-Aug	1.32	−2.74	1	1, 0.1
7-Sep	2.36	−1.90	0.98	1, 0.1, 0.05

[0223]

TABLE 10

Characteristic melting temperatures (T_m ± CV %) of the microcystin synthetase E quantitative real-time PCR amplification products (247 bp) obtained using LightCycler melting curve analysis. Nucleotide differences were calculated for the 209 bp long sequence between the primer annealing sites. Number of samples is denoted by n.

			Nucleotide differences					
			Anabaena			*Microcystis*		
Strain	T_m ± CV %	n	90	315	202 A1	GL 26 0735	PCC 7806	PCC 7941
Anabaena								
90	79.7 ± 0.2	12						
315	79.3 ± 0.4	14	0					
202A1	79.7 ± 0.2	12	1	1				

TABLE 10-continued

Characteristic melting temperatures (T_m ± CV %) of the microcystin synthetase E quantitative real-time PCR amplification products (247 bp) obtained using LightCycler melting curve analysis. Nucleotide differences were calculated for the 209 bp long sequence between the primer annealing sites. Number of samples is denoted by n.

			Nucleotide differences					
			Anabaena			*Microcystis*		
Strain	T_m ± CV %	n	90	315	202 A1	GL 26 0735	PCC 7806	PCC 7941
Microcystis								
GL 260735	81.3 ± 0.2	12	45	45	46			
PCC 7806	81.5 ± 0.2	15	47	47	48	2		
PCC 7941	81.5 ± 0.1	11	47	47	48	2	1	

Example 4

[0224] We were interested in the mcyD gene region as part of an evolutionary study on microcystin synthetase genes from different genera of cyanobacteria.

[0225] The McyD gene is involved in the formation of the Adda amino acid and this amino acid along with D-glutamate is critical to microcystin toxicity (Goldberg, J., Huang, H-B., Kwon, is Y-G., Greengard, P., Nairn, A. C. et al. Three-dimensional structure of the catalytic subunit of protein serine/threonine phosphatase-1. Nature 376, 745-753 (1995). The Adda amino acid is proposed to be assembled by McyG, McyD and mcyE (Tillett, D. et al. Structural organization of microcystin biosynthesis in *Microcystis aeruginosa* PCC7806: an integrated peptide-polyketide synthetase system. Chem. Biol. 7, 753-764 (2000). The mcyD gene region we sequenced encodes parts of a beta-ketoacyl synthase and a acyltransferase domain (Tillett et al. 2000). The region we looked at is specifically involved in one round of chain elongation of the growing Adda amino acid (Tillett et al. 2000).

[0226] The 818 bp region of the mcyD gene was amplified with the mcyDF (5'-gatccgattgaattagaaag-3) and mcyDR (5'-gtattccccaagattgcc-31) primers. PCR reactions for the mcyD PCR products were performed in a 20 ml final volume containing 1 ml of DNA, 1×DynaZyme II PCR buffer, 250 mM of each deoxynucleotide, 0.5 mM of both PCR primers, and 0.5 U of DynaZyme II DNA polymerase (Finnzymes, Espoo, Finland). The following thermocycle protocol was used: 95° C., 3 min; 30×(94° C., 30 sec; 56° C., 30 sec; 72° C., 1 min); 72° C., 10 min. Sequencing of the mcyD PCR products was performed by Genome Express (France).

Example 5

Oligonucleotides for Detection and Identification of Toxic Cyanobacteria

Materials and Methods

[0227] All chemicals and solvents were purchased from Sigma-Aldrich (Italy) and used without further purification. Oligonucleotides were purchased from Interactiva Biotechnologie GmbH (Germany).

DNA Samples

[0228] The samples used to validate the probes were *Anabaena* 202A1, *Microcystis* 205, *Planktothrix* 49, *Nostoc* 152 and the environmental samples 0TU35 (>10 um fraction) and 0TU33 (bloom sample).

Ligation Probe Design

[0229] For Ligation Detection Reaction, we designed specific probes for the mcyE sequences of five different groups. These groups were identified using a phylogenetic tree obtained from the ARB software, version Beta 011107.

[0230] ARB (www.arb-home.de) is a UNIX-based program for aligning a large number of DNA sequences and for constructing phylogenetic trees according to a central database of processed sequences.

[0231] The mcyE sequences were aligned using CLUSTAL W (Thompson et al., 1994) and internal ARB algorithms. The phylogenetic tree was constructed using the neighbor-joining (NJ) algorithm (Saitou and Nei, 1987). The groups are the following: *Anabaena, Microcystis, Nodularia, Nostoc, Oscillatoria/Planktothrix* (OP).

[0232] From the sequence alignment a "group-specific" consensus sequence was obtained with a cutoff percentage of 95%. This value is compared with the frequency of the residues found at each alignment position. If the residue at a given position occurred at a lower frequency than the cutoff percentage, an IUPAC ambiguous symbol was displayed in the consensus sequence.

[0233] Then, group-specific probe design was obtained using a tool on ARB database named "Probe design".

[0234] All oligonucleotides were designed to have a melting temperature (T_m) between 64 and 68° C.

[0235] Discriminating probes were purchased with a Cy3 label at their 5' terminal position and common probes with a phosphate in the same position.

Universal Array Preparation

[0236] Microarrays were prepared using CodeLink™ slides (Amersham Biosciences), designed to covalently immobilize NH_2-modified oligonucleotides.

[0237] 5' amino-modified Zip Code oligonucleotides, carrying an additional poly$(dA)_{10}$ tail at their 5' end, were

diluted to 25 μM in 100 mM phosphate buffer (pH 8.5). Spotting was performed using a non contact piezo driven dispensing system (Nanoplotter, GeSim, Germany). Printed slides were processed according to the manufacturer's protocols.

[0238] Quality control of printed surfaces was performed by sampling one slide from each deposition batch. The printed slide was hybridized with 1 μM 5' Cy3 labeled poly(dT)$_{10}$ in a solution containing 5×SSC and 0.1 mg/ml salmon sperm DNA at RT for 2 h, then washed for 15 min in 1×SSC. The fluorescent signal was controlled by laser scanning following procedures described in "Array hybridization, detection and data analysis".

[0239] PCR Amplifications from DNA Samples.

Ligation Detection Reaction.

[0240] Ligation Detection Reaction was carried out in a final volume of 20 μl containing 20 mM Tris-HCl (pH 7.5), 20 mM KCl, 10 mM $MgCl_2$, 0.1% NP40, 0.01 mM ATP, 1 mM DTT, 2 pmol of each discriminating probe, 2 pmol of each common probe and 100 fmol of purified PCR products. The reaction mixture was preheated for 2 min at 94° C. and spinned in a microcentrifuge for 1 min; then 1 ul of 4 U/ul Pfu DNA ligase (Stratagene, La Jolla, Calif.) was added. Alternatively, 0.5 ul of 50 U/ul Tth DNA ligase (ABgene) was used.

[0241] The LDR was cycled for 30 rounds of 90° C. for 30 sec and 60° C. for 4 min in the GeneAmp PCR system 9700 thermal cycler (Applied Biosystems, California).

Array Hybridization, Detection and Data Analysis.

[0242] In a 0.5-ml microcentrifuge tube, the LDR mix (20 μl) was diluted to obtain 65 μl of hybridization mixture containing 5×SSC and 0.1 mg/ml salmon sperm DNA. The mix, after heating at 94° C. for 2 min and chilling on ice, was applied onto the slide under a hybridization chamber.

[0243] Hybridization was carried out in the dark at 65° C. for two hours in a temperature-controlled water bath. After hybridization, the microarray was washed at 65° C. for 15 min in pre-warmed 1×SSC, 0.1% SDS. Finally, the slide was spinned at 80 g for 3 min.

[0244] The fluorescent signals were acquired at 5 μm resolution using a ScanArray® 4000 laser scanning system (PerkinElmer Life Sciences) with green laser for Cy3 dye (λ_{ex} 543 nm/λ_{em} 570 nm). Both the laser and the photomultiplier (PMT) tube power were set at 70-95%. To quantitate the fluorescent intensity of the spots we used the QuantArray Quantitative Microarray Analysis software (Perkin Elmer Life Sciences).

[0245] Recently, we have presented a Universal DNA Array approach to discriminate some groups of bacteria (Busti et al., 2002). This procedure, based on the discriminative properties of the DNA ligation reaction, requires the design of two probes specific for each target sequence, as described by Barany and co-workers (1999). One oligonucleotide brings a fluorescent label and the other a unique sequence named complementary Zip Code (cZip Code). Ligated fragments, obtained in presence of a proper template by the action of a DNA ligase, are addressed to the location on the microarray where the Zip Code sequence has been spotted. Such an array is therefore "Universal" being unrelated to a specific molecular analysis.

[0246] Here we present the Universal DNA Array approach applied to the detection of cyanobacterial mcyE gene diversity.

Ligation Probes Design

[0247] We used the ARB software to perform the sequence alignment of cyanobacterial mcyE sequences. These sequences were aligned and clustered according to their phylogenetic lineages so that 5 "group-specific" consensus sequences were yielded: *Anabaena, Microcystis, Nodularia, Nostoc, Oscillatoria/Planktothrix* (OP) (FIG. **14**). Then, "group-specific" probes were designed using a tool on ARB database named "Probe design". Among this set of probes, we selected discriminating probes with 3' position unique to each group in order to obtain ligase discrimination. As a matter of fact, after hybridization of a discriminating probe and a common probe to the target sequence, ligation occurs only if there is perfect complementarity at the junction between the two oligos. Common probes were designed immediately 3' to the discriminating oligo from the group-specific consensus.

[0248] All the selected probes are described in FIG. **20**. We selected one probe pair for each group of interest, except for the *Oscillatonia/Planktothrix* group.

[0249] FIG. **15** shows the alignment of the "group-specific" consensus sequences and the relative discriminating probes.

Zip Codes Assignment and Quality Control of the Universal Array

[0250] We randomly selected 6 Zip code sequences from those described by Chen and co-workers, 2000. Each Zip code was randomly assigned to a single cyanobacterial group. Each common probe was synthesized to have the complementary Zip code (cZip code) affixed to its 3' end (FIG. **20**). No significant self-annealing of the common probe-cZip sequences was detected by computer analysis (data not shown).

[0251] The Zip codes were deposited using a non contact deposition system. The deposition scheme is shown in FIG. **17**. In order to verify the deposition quality of the Zip Code oligonucleotides on the slides, we performed hybridisations with Cy3 labelled poly(dT) complementary to the poly(da)$_{10}$ sequence of each Zip Code. Every controlled slide revealed intense fluorescent signals corresponding the spotted oligonucleotides, as shown in FIG. **17**.

[0252] This result indicated a rather uniform deposition of the oligos on the Universal Array.

LRD detection onto Universal Array

1) Probes Specificity

[0253] The specificity of the probes for mcyE cyanobacterial groups was tested using PCR amplified fragment of this gene coming either from pure strains or from environmental samples, as indicated in Materials and Methods.

[0254] LDRs were conducted in the presence of the PCR product of each single sample as template and in the presence of all the probes (discriminating probes and common probes).

[0255] A negative control of the entire process was performed using double distilled water instead of genomic. DNA as PCR substrate. After standard cycling, ten microliters of the reaction mixture were used in the LDR. Following hybridisation on the universal chip, no signal was detected even setting PMT and laser to 95% of their power (data not shown).

[0256] In the presence of the proper DNA template, the Universal Array behaved as expected: only group-specific spots showed positive signal. The results are showed in FIG. **18**.

2) Probe Sensitivity

[0257] In order to establish the detection limit of the method, we performed the Ligation Detection Reaction starting from 50, 5 and 1 fmol of three different PCR products as substrates. The detected signals progressively decrease and three visible signals were detected up to 1 finol of the PCR products. No signals were detected using 0.5 fmol of the substrates even setting PMT and laser to 95% of their power (data not shown).

Example 6

Molecular Analysis of Cyanobacterial Diversity by Microarrays on "PCR-Amplified" 16 rRNA Gene

[0258] All chemicals and solvents were purchased from Sigma-Aldrich (Italy) and used without further purification. Oligonucleotides were purchased from Interactiva Biotechnologie GmbH (Germany).

DNA Samples

[0259] The samples used to validate the probes included axenic strains kept in the authors' culture collections, strains isolated from European lakes and a reservoir during this study, and clones of environmental DNA libraries obtained from Lake Esch-sur-Sûre (Luxembourg) and Lake Tuusulanjärvi (Finland). The 16S rRNA gene of the cultured strains and clones was sequenced (unpublished data). In addition, the array was tested with an environmental DNA sample (Lake Tuusulanjärvi), which was isolated with the hot-phenol method. To verify the microarray results, the same environmental sample was analyzed with DGGE and cloning of the 16S rRNA gene.

Ligation Probe Design

[0260] For Ligation Detection Reaction, we designed specific probes for the 16S rRNA gene sequences of different cyanobacterial groups. These groups were identified using a cyanobacterial 16S rRNA gene tree obtained from the ARB software, version Beta 011107.

[0261] ARB (www.arb-home.de) is a UNI-based program for aligning a large number of 16S rRNA gene sequences and for constructing phylogenetic trees according to a central database of processed sequences. ARB cyanobacterial 16S rDNA database we used contained 281 sequences from public databases and 57 from this study, in addition to the outgroup *Escherichia coli*. All these sequences were longer than 1400 bp, except the two sequences of Antarctic *Phormidium* (about 1350 bp) and 21 (out of 42) sequences of *Prochlorococcus marinus* (about 1250 bp). All sequences were aligned with CLUSTAL W (24) and ARB. The phylogenetic tree was constructed using the neighbor-joining (NJ) algorithm (Saitou and Nei, 1987). As shown in FIG. **25**, the selected cyanobacterial groups are the following: *Anabaena/Aphanizomenon, Calothrix, Cylindrospermopsis, Cylindrospermum, Gloeothece, Halotolerants, Leptolyngbya, Palau Lyngbya, Microcystis, Nodularia, Nostoc, Oscillatoria/Planktothrix, Antarctic Phormidium, Prochlorococcus, Spirulina, Synechococcus, Synechocystis, Trichodesmium, Woronichinia.*

[0262] From the sequence alignment a "group-specific" consensus sequence was obtained with a cutoff percentage of 75%. This value is compared with the frequency of the residues found at each alignment position. If the residue at a given position occurred at a lower frequency than the cutoff percentage, an IUPAC ambiguous symbol was displayed in the consensus sequence.

[0263] Then, the 19 group consensus sequences were imported in GCG Omiga 2.0 (Oxford Molecular Ltd.) for group-specific probe design. The specificity of each probe pair (discriminating probe and common probe) was controlled on the entire bacterial 16S rDNA ARB database. All oligonucleotides were designed to have a melting temperature (T_m) between 64 and 68° C.

[0264] Discriminating probes were purchased with a Cy3 label at their 5' terminal position and common probes with a phosphate in the same position.

Universal Array Preparation

[0265] Microarrays were prepared using CodeLink™ slides (Amersham), designed to covalently immobilize NH_2-modified oligonucleotides.

[0266] 5' amino-modified Zip Code oligonucleotides, carrying an additional poly$(dA)_{10}$ tail at their 5' end, were diluted to 25 μM in 100 mM phosphate buffer (pH 8.5). Spotting was performed using a contact dispensing system MicroGrid II (BioRobotics). Printed slides were processed according to the manufacturer's protocols. 8 subarrays per slide were generated.

[0267] Quality control of printed surfaces was performed by sampling one slide from each deposition batch. The printed slide was hybridized with 1 μM 5' Cy3 labeled poly$(dT)_{10}$ in a solution containing 5×SSC and 0.1 mg/ml salmon sperm DNA at RT for 2 h, then washed for 15 min in 1×SSC. The fluorescent signal was controlled by laser scanning following procedures described in "Array hybridization, detection and data analysis".

[0268] PCR Amplifications from DNA Samples.

[0269] The DNA region coding for 16S ribosomal RNA was amplified with a universal primer 16SF27 (5'AGAGMTIGATCMTGGCTCAG 3') (Edwards et al., 1989) and a cyanobacterial specific primer 23S30R (5'CCTCGCCTCTGTGTGCCTAGGT3) (Lepère et al., 2000) which permitted the amplification of a ca 2000 bp fragment.

[0270] PCR amplifications were performed in a GeneAmp PCR system 9700 thermal cycler (Applied Biosystem, California). The reaction mixtures include 500 nM each primer, 200 μM each dNTP, 10 mM Tris-HCl (pH 8.8), 1.5 mM MgC_2, 50 mM KCl, 0.1% (wt/vol) Triton X-100, 1 U of DynaZyme DNA polymerase (Finnzymes OY, Espoo, Finland) and 5-8 ng of genomic DNA, in a final volume of 50 μl. Prior to amplification, DNA was denatured for 5 min at

95° C. Amplification consisted of 30 cycles of 94° C. for 45 s, 57° C. for 45 s and 72° C. for 2 min. After the cycles, an extension step (10 min at 72° C.) was performed.

[0271] The PCR products were purified by GFX PCR DNA purification kit (Amersham Biosciences, Piscataway-NJ), eluted in 50 μl of autoclaved water and quantified by the BioAnalyzer 2100 (Agilent Technologies).

Ligation Detection Reaction

[0272] Ligation Detection Reaction was carried out in a final volume of 20 μl containing 20 mM Tris-HCl (pH 7.5), 20 mM KCl, 10 mM $MgCl_2$, 0.1% NP40, 0.01 mM ATP, 1 mM DTT, 250 fmol of each discriminating probe, 250 fmol of each common probe, 10 fmol of the hybridization control and 25 fmol of purified PCR products. The reaction mixture was preheated for 2 min at 94° C. and spinned in a microcentrifuge for 1 min; then 1 ul of 4 U/ul Pfu DNA ligase (Stratagene, La Jolla, Calif.) was added. The LDR was cycled for 30 rounds of 90° C. for 30 sec and 60° C. for 4 min in the GeneAmp PCR system 9700 thermal cycler (Applied Biosystems, California).

Array Hybridization, Detection and Data Analysis.

[0273] In a 0.5-ml microcentrifuge tube, the IDR mix (20 μl) was diluted to obtain 65 μl of hybridization mixture containing 5×SSC and 0.1 mg/ml salmon sperm DNA. The mix, after heating at 94° C. for 2 min and chilling on ice, was applied onto the slide in the Press-To-Seal Silicone Isolators 1.0×9 mm (Schleicher & Schuell).

[0274] Hybridization was carried out in a hybridization chamber in the dark at 65° C. for two hours in a temperature-controlled water bath. After hybridization, the microarray was washed at 65° C. for 15 min in pre-warmed 1×SSC, 0.1% SDS. Finally, the slide was spinned at 80 g for 3 min.

[0275] The fluorescent signals were acquired at 5 μm resolution using a ScanArray® 4000 laser scanning system (PerkinElmer Life Sciences) with green laser for Cy3 dye (λ_{wl} 543 nm/λ_{em} 570 nm). Both the laser and the photomultiplier (PMT) tube power were set at 70-95%.

[0276] To quantify the fluorescent intensity of the spots we used the QuantArray Quantitative Microarray Analysis software (Perkin Elmer Life Sciences).

[0277] When statistical analyses were performed, we included the fluorescent intensity values obtained from replicated spots (four replicates spot for each group, eight replicates spot for the universal) and replicates experiments sets (three LDR-universal array experiments).

Sequence Analysis of Cyanobacterial 16S rDNA and Ligation Probes Design

[0278] We used the ARB software to perform the sequence alignment of cyanobacterial 16S rDNA. The ARB database we used contained 281 cyanobacterial sequences from public databases and 57 from this study. These sequences were aligned and clustered according to their phylogenetic lineages so that 19 "group-specific" consensus sequences were yielded (FIG. **25**).

[0279] Then, the 19 group consensi were imported in GCG Omiga 2.0 (Oxford Molecular Ltd.). The Omiga software is a graphically oriented package that permits the identification of "group-specific" nucleotide polymorphisms. Thus, the probes were designed complementary to polymorphic regions on the basis of a final alignment among group-specific consensi. The selection process consisted in several steps. Firstly, we considered the ligase reaction features. As shown in FIG. **26**, after hybridization of a discriminating probe and a common probe to the target sequence, ligation occurs only if there is perfect complementarity at the junction between the two oligos. For this reason, to obtain ligase discrimination, we selected discriminating probes with 3' position unique to each group. Common probes were designed immediately 3' to the discriminating oligo from the group-specific consensus.

[0280] Secondly, among this set of probes, we selected only those pairs of probes, which differed from all representatives of the other groups at least for the 3' terminal position of the discriminating probes, but which were invariant in all members of their group. Examples of probe design procedure are shown in FIG. **27**.

[0281] Finally, in order to discard potentially a specific probe pairs, we analyzed each probe pair (discriminating probe and common probe) using a tool on ARB database, which permit to verify probes against all the bacterial 16S rRNA gene sequences. Initially, we considered 60 group specific probe pairs, but only 21 of these have been chosen after the selection step described above.

[0282] All the selected probes are described in FIG. **32**. When the consensus sequence contains a degenerate base, we included inosine during oligonucleotide synthesis at these degenerate positions.

[0283] Although DNA samples for some of the 19 selected groups (i.e. *Gloeothece, Antarctic Phormidium, Prochlorococcus marinus, Trichodesmium*) were not available because these cyanobacteria are not present in the lakes under scrutiny, all the ARB phylogenetic lineages have been considered in the experimental set-up to allow for future applications of this cyanobacterial microarray.

[0284] In order to have a positive control for the Ligation Detection Reaction, a universal probe pair, matching all the cyanobacteria, was designed and the corresponding Zip code was included in the Universal Array. As a positive control for the hybridisation reaction, a Cy3 labelled complementary Zip Code sequence was added in the hybridization mixture and the corresponding Zip code was included in the Universal Array.

Zip Codes Assignment and Quality Control of the Universal Array

[0285] We randomly selected 21 Zip code sequences from those described by Barany and coworkers and Chen and co-workers. Each Zip code was randomly assigned to a single cyanobacterial group, except Zip code1 which is the positive control for the hybridisation reaction.

[0286] Each common probe was synthesized to have the complementary Zip code (cZip code) affixed to its 3' end (FIGS. **32** and **39**). No significant self-annealing of the twenty common probe-cZip sequences was detected by computer analysis (data not shown).

[0287] The Zip codes were deposited using a contact deposition system generating 8 subarrays per slide. The deposition scheme is shown in FIG. **28**. In order to verify the deposition quality of the Zip Code oligonucleotides on the

slides, we performed hybridisations with Cy3 labelled poly(dt) complementary to the poly(da)$_{10}$ sequence of each Zip Code.

LDR Detection onto Universal Array of Cyanobacterial 16S rDNA Samples

1) Probes Specificity

[0288] The specificity of the probes for freshwater cyanobacterial groups was tested using PCR amplified 16S rRNA gene coming either from pure strains (both axenic and isolated in this study) or from cloned rDNA sequences. All pure strains used to validate the LDR probes are described in FIG. **33**. The sequences obtained from the clones have been aligned in the ARB database with the sequences of pure cyanobacterial strains in order to define their phylogenetic group. The clones used are described in FIG. **34**.

[0289] LDRs were conducted in the presence of the PCR product of each single strain or clone as template and in the presence of all the probes (discriminating probes and common probes).

[0290] A negative control of the entire process was performed using double distilled water instead of genomic DNA as PCR template. After standard cycling, ten microliters of the reaction mixture were used in the LDR. Following hybridisation on the Universal Array, no signal was detected even setting PMT and laser to 95% of their power (data not shown).

[0291] In the presence of the proper DNA template, the Universal Array behaved as expected: only group specific spots, universal spots and the spots corresponding to the hybridization control showed positive signal. Some of the results are shown in FIG. **29**.

2) Probe Sensitivity

[0292] In order to establish the detection limit of the method and the correlation between signal intensity and template concentration, we performed Ligation Detection Reactions starting from 100 to 0,5 fmol of PCR products obtained from Planktothrix 1LT as substrates. The detected signals progressively decrease and a visible signal was detected up to 1 fmol of the PCR product. No signals were detected using 0.5 fmol of the substrates even setting PMT and laser to 95% of their power (data not shown). The linear correlation between signal intensity and template concentration is shown in FIG. **31**.

3) Use of Artificial Mixes of PCR Products from Different Strains.

[0293] In order to determine the efficiency of the LDR method in presence of complex molecular targets, we used artificial mixes with unbalanced amounts of PCR products derived from the following cyanobacterial samples: *Aphanizomenon* sp. 202, *Microcystis* OBB 34S, *Spirulina subsalsa* PCC6313, *Calotlrix* sp. PCC7714, clone Woronichinia OES46. After separate PCR reactions, the amplified fragments were pooled in unbalanced LDR mixes using different ratios: 100:1, 50:1, 100:5, 50:5. In all these experiments *Aphanizomenon* sp. 202 and *Microcystis* OBB 34S were the more concentrated samples. Moreover, we mixed also 500 fmol of the amplicon derived from *Microcystis* OBB 34S with 5 fmol of the PCR fragment obtained from Woronichinia OES46 clone. After the hybridization of the LDR products onto the Universal Array, the signals related to the lower concentrated template were not detected in the LDR mixes with these ratio: 100:1 and 50:1. Only in presence of the LDR products obtained from the mixes with the ratio 100:5 and 50:5 all the expected signals are detected FIG. **29**. The fluorescent intensity of the spots was quantified and the results are shown in FIG. **29**. Furthermore, we compared also the results obtained using two LDR unbalanced mixes 100:1 (100 fmol of *Microcystis* OBB 34S and 1 fmol each of Spirina, Woronichinia and Calothrix), in one of which 8 U of Pfa DNA ligase was added, whereas the other was prepared using 4 U of the enzyme, as described in Materials and Methods. Hybridization signals of the lower concentrated substrates were detected only from the LDR product obtained using 8 U of Pfu DNA ligase instead of 4 U (FIG. **30**).

[0294] LDR Detection onto Universal Array of 16S rDNA and mcyE from Environmental Samples

[0295] We made PCR amplification from genomic DNA using 16S cyanobacteria specific primers. The PCR conditions used are shown in FIG. **35**. We made also PCR amplification from genomic DNA using mcyE gene primers. The ligation detection reaction was made under the same conditions by using an oligo mix containing both the probes for 16S rRNA gene and the probes for the mcyE gene as shown in FIG. **36**. Finally the hybridization was carried on the same Universal Array where the 16S rRNA LDR product and, mcyE LDR product were detected

Example 7

Microarray Platform for Toxic and Non-Toxic Detection in Cyanobacteria.

Materials and Methods.

[0296] All chemicals and solvents were purchased from Sigma-Aldrich (Italy) and used without further purification. Oligonucleotides were purchased from Interactiva Biotechnologie GmbH (Germany).

Ligation Probe Design

[0297] The mcyE probe design has been previously described in Example 5 in "Ligation probe design". The 16S rRNA gene probe design has been previously described in Example 6 in "Ligation probe design", but was added the probe design for a further cyanobacteria group: Snowella. The Snowella probe design was performed using the updated ARB database containing 281 sequences from public databases and 69 from this study (FIG. **25**B). The updated database allowed to design specific probe for *Aphanizomenon* and *Anabaena* subgroups as shown in FIG. **25**C. The probe design allows the detection of 20 toxic and non-toxic cyanobacteria groups.

Universal Array Preparation

[0298] Microarrays were prepared using CodeLink™ slides (Amersham), designed to covalently immobilize NH_2-modified oligonucleotides.

[0299] 5' amino-modified Zip Code oligonucleotides, carrying an additional poly(daA)$_{10}$ tail at their 5' end, were diluted to 25 M in 100 mM phosphate buffer (pH 8.5). Spotting was performed using a contact dispensing system

MicroGrid II (BioRobotics). Printed slides were processed according to the manufacturer's protocols. 8 subarrays per slide were generated.

[0300] The Universal array used for the detection of toxic and non-toxic cyanobacteria was designed to detect both the 16 rRNA and mcyE gene ligated probes. For this purpose the deposition scheme was improved as shown in FIG. **27**B. We generated 8 subarray per slide. Each subarray is made of 208 spots including zipcodes for hybridization control, cyanobacterial universal probes, 16S rRNA gene specific probe, mcyE specific probe and empty spot as a negative control. Each specific zip code for the recognition of cyanobacteria universal probe, 16Ss RNA gene probe and mcyE gene probe is spotted in quadruplicate. The LDR positive control (zipcode no 63) is replicated 6 times, while the hybridization positive control (zipcode no 66) is replicated 8 times.

[0301] Quality control of printed surfaces was performed by sampling one slide from each deposition batch. The printed slide was hybridized with 1 μM 5' Cy3 labeled poly(dt)$_{10}$ in a solution containing 5×SSC and 0.1 mg/ml salmon sperm DNA at RT for 2 h, then washed for 15 min in 1×SSC. The fluorescent signal was controlled by laser scanning following procedures described in "Array hybridization, detection and data analysis".

PCR Amplification from DNA Samples

[0302] The PCR of mcyE gene and 16S rRNA gene were performed separately, using the conditions previously described in Examples 5 and 6 in "PCR amplification from DNA samples".

Ligation Detection Reaction

[0303] The Ligation Detection Reaction for toxic and non-toxic cyanobacteria detection was done mixing together the PCR product of 16S rRNA and mcyE gene and the discrimination and common probe specific for both 16s rRNA and mcyE gene, FIG. **36**.

[0304] Ligation Detection Reaction was carried out in a final volume of 20 μl containing 20 mM Tris-HCl (pH 7.5), 20 mM KCl, 10 mM $MgCl_2$, 0.1% NP40, 0.01 mM ATP, 1 mM DTT, 250 fmol of each discriminating probe, 250 fmol of each common probe, 10 fmol of the hybridization control and 25 fmol of purified PCR products. The reaction mixture was preheated for 2 min at 94° C. and spinned in a microcentrifuge for 1 min; then 1 ul of 4 U/ul Pfu DNA ligase (Stratagene, La Jolla, Calif.) was added. The LDR was cycled for 30 rounds of 90° C. for 30 sec and 60° C. for 4 min in the GeneAmp PCR system 9700 thermal cycler (Applied Biosystems, California).

Array Hybridization, Detection and Data Analysis

[0305] In a 0.5-ml microcentrifuge tube, the LDR mix (20 μl) was diluted to obtain 65 μl of hybridization mixture containing 5×SSC and 0.1 mg/ml salmon sperm DNA. The mix, after heating at 94° C. for 2 min and chilling on ice, was applied onto the slide in the Press-To-Seal Silicone Isolators 1.0×9 mm (Schleicher & Schuell).

[0306] Hybridization was carried out in a hybridization chamber in the dark at 65° C. for two hours in a temperature-controlled water bath. After hybridization, the microarray was washed at 65° C. for 15 min in pre-warmed 1×SSC, 0.1% SDS. Finally, the slide was spinned at 80 g for 3 min.

[0307] The fluorescent signals were acquired at 5 μAm resolution using a ScanArray® 4000 laser scanning system (PerkinElner Life Sciences) with green laser for Cy3 dye (λ_{ex} 543 nm/λ_{em} 570 nm). Both the laser and the photomultiplier (PMT) tube power were set at 70-95%.

[0308] To quantify the fluorescent intensity of the spots we used the QuantArray Quantitative Microarray Analysis software (Perkin Elmer Life Sciences).

[0309] When statistical analyses were performed, we included the fluorescent intensity values obtained from replicated spots (four replicates spot for each group, eight replicates spot for the universal) and replicates experiments sets (three LDR-universal array experiments).

Zip Codes Assignment and Quality Control of the Universal Array

[0310] We randomly selected 33 Zip code sequences from those described by Chen and co-workers, 2000. Each Zip code was randomly assigned to a single cyanobacterial group. Each common probe, for both 16S rRNA and mcyE gene recognitin, was synthesized to have the complementary Zip code (cZip code) affixed to its 3' end (FIGS. **20**, **32** and **39**). No significant self-annealing of the common probe-cZip sequences was detected by computer analysis (data not shown).

[0311] The Zip codes were deposited using a contact deposition system. The deposition scheme is shown in FIG. **27**B. In order to verify the deposition quality of the Zip Code oligonucleotides on the slides, we performed hybridisations with Cy3 labelled poly(dT) complementary to the poly(da)$_{10}$ sequence of each Zip Code. Every controlled slide revealed intense fluorescent signals corresponding the spotted oligonucleotides, as shown in FIG. **27**B. This result indicated a rather uniform deposition of the oligos on the Universal Array.

LDR Detection onto Universal Array of Cyanobacterial 16S rDNA and mcyE Samples

Probes Specificity

[0312] The specificity of the probes was tested using PCR amplified 16S rRNA and mcyE gene coming from pure strains (both axenic and isolated in this study.)

[0313] A negative control of the entire process was performed using double distilled water instead of genom DNA as PCR template. After standard cycling, ten microliters of the reaction mixture were used in the LDR. Following hybridisation on the Universal Array, no signal was detected even setting PMT and laser to 95% of their power (data not shown).

[0314] In the presence of the proper DNA template of both 16S rRNA and mcyE genes, the Universal Array functioned very well: only group specific spots, universal spots and the spots corresponding to the hybridization control showed positive signal. Some of the results are shown in FIG. **30**B.

REFERENCES

[0315] Altschul, S. F., T. L. Madden, A. A Schäffer, J. Zhang, Z. Zhang, W. Miller, and D. J. Lipman. 1997. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic. Acids. Res. 25:3389-3402.

[0316] Amard, B. & Bertrand-Sarfati, J. Microfossils in 2000 Ma old cherty stromatolites of the Franceville Group, Gabon. *Precambrian Res.* 81' 197-221 (1997).

[0317] Aparicio, J. F., Molnár, I., Schwecke, T., König, A., Haydock, S. F., Khaw, L. E., Staunton, J., Leadlay, P. F., 1996. Organization of the biosynthetic gene cluster for rapamycin in *Streptomyces hygroscopicus*: analysis of the enzymatic domains in the modular polyketide synthase. Gene 169, 9-16.

[0318] Baker, J. A., B. A Neilan, B. Entsch, and D. B. McKay. 2001. Identification of cyanobacteria and their toxigenicity in environmental samples by rapid molecular analysis. Environ. Toxicol. 16:472-482.

[0319] Baker, J. A., B. Entsch, B. A Neilan, and D. B. McKay. 2002. Monitoring changing toxigenicity of a cyanobacterial bloom by molecular methods. Appl. Environ. Microbiol. 68:6070-6076.

[0320] Becker, S., M. Fahrbach, P. Böger, and A. Ernst. 2002. Quantitative tracing, by Taq nuclease assays, of a *Synechococcus* ecotype in highly diversified natural population. Appl. Environ. Microbiol. 68:4486-4494.

[0321] Becker, S., P. Böger, R. Oehlmann, and A Ernst. 2000. PCR bias in ecological analysis: a case study for quantitative Taq nuclease assays in analyses of microbial communities. Appl. Environ. Microbiol. 66:4945-4953.

[0322] Boers, P., L. van Ballegooijen, and J. Uunk. 1991. Changes in phosphorus cycling in a shallow lake due to food web manipulations. Freshwater Biol. 25:9-20.

[0323] Brocks, J. J., Logan, G. A., Buick, R., & Summons, R. E. Archean molecular fossils and the early rise of eukaryotes. Science 285, 1033-1036 (1999).

[0324] Busti E, Bordoni R, Castiglioni B, Monciardini P, Sosio M, Donadio S, Consolandi C, Rossi Bernardi L, Battaglia C, De Bellis G. Bacterial discrimination by means of a universal array approach mediated by LDR (ligase detection reaction). BMC Microbiol 2:27 (2002).

[0325] Castenholz, R. W. 2001. Phylum BX. Cyanobacteria, oxygenic photosynthetic bacteria. p. 473-599. In D. R. Boone, R. W. Castenholz and G. M. Garrity (ed.), Bergey's manual of systematic bacteriology, 2nd edition, vol. 1. Springer-Verlag.

[0326] Cavalier-Smith T. Origins of secondary metabolism. *Ciba Found Symp.* 171, 64-80 (1992).

[0327] Chen, J. et al. A microsphere-based assay for multiplexed single nucleotide polymorphism analysis using single base chain extension. Genome Res. 10, 549-557 (2000). A clear presentation of the concept of generic 'tag' sequences as applied to SNP genotyping.

[0328] Chorus I., and L. Mur. 1999. Preventative measures. p. 235-273. In I. Chorus and J. Bartram (ed.), Toxic cyanobacteria in water. E & FN Spon. London and New York.

[0329] Christiansen, G., Fastner, J., Erhard, M., Börner, T., Dittmann, E., 2003. Microcystin Biosynthesis in *Planktothrix*: Genes, Evolution, and Manipulation. J. Bacteriol. 185, 564-572.

[0330] DeMott, W. R. & Moxter, F. Foraging on cyanobacteria by copepods: responses to chemical defenses and resource abundance. *Ecology* 72, 1820-1834 (1991).

[0331] Dittmann, E., Neilan, B. A., Erhard, M., von Döhren, H., Börner, T., 1997. Insertional mutagenesis of a peptide synthetase gene that is responsible for hepatotoxin production in the cyanobacterium *Microcystis aeruginosa* PCC7806. Mol. Microbiol. 26, 779-787.

[0332] Dittmann, E. et al. Altered expression of two light-dependent genes in a microcystin-lacking mutant of *Microcystis aeruginosa* PCC 7806. Microbiol. 147, 3113-3119 (2001).

[0333] Du, L., Sanchez, C., Chen, M., Edwards, D. J., Shen, B., 2000. The biosynthetic gene cluster for the antitumor drug bleomycin from *Streptomyces verticillus* ATCC15003 supporting functional interactions between nonribosomal peptide synthesis and a polyketide synthase. Chem. Biol. 7, 623-642.

[0334] Duitman, E. H., Hamoen, L. W., Rembold, M., Venema, G., Seitz, H., Saenger, W., Bernhard, F., Reinhardt, R., Schmidt, M., Ullrich, C., Stein, T., Leenders, F., Vater, J., 1999. The mycosubtilin synthetase of *Bacillus subtilis* ATCC6633: A multifunctional hybrid between a peptide synthetase, an amino transferase, and a fatty acid synthase. Proc Natl Acad Sci USA 96, 13294-13299.

[0335] Edwards, U., Rogall, T., Blöcker, H., Emde, M. & Böttger, E. C. Isolation and direct complete nucleotide determination of entire genes. Characterization of a gene coding for 16S ribosomal RNA. *Nucleic Acids Res.* 17, 7843-7853 (1989).

[0336] Ekman-Ekebom, M., M. Kauppi, K Sivonen, M. Niemi, and L. Lepistö. 1992. Toxic cyanobacteria in some Finnish lakes. Environ. Toxicol. Wat. Qual. 7:201-213.

[0337] Falconer, I., J. Bartram, I. Chorus, T. Kuiper-Goodman, H. Utkilen, M. Burch, and G. A. Codd. 1999. Safe levels and safe practices. p. 155-178. In I. Chorus and J. Bartram (ed.), Toxic cyanobacteria in water. E & FN Spon. London and New York.

[0338] Fastner, J., Erhard, M., von Döhren H., 2001. Determination of oligopeptide diversity within a natural population of *Microcystis* spp. (cyanobacteria) by typing single colonies by matrix-assisted laser desorption ionization-time of flight mass spectrometry. Appl. Environ. Microbiol. 67, 5069-5076.

[0339] Felsenstein, J. *PHYLIP* (*Phylogenetic Inference Package*), version 3.5c. (Department of Genetics, Univ. Washington, Seattle, 1993).

[0340] Fujii, K, Harada, K, Suzuki, M., Kondo, F., Ikai, Y., Oka, H., Carmichael, W. W., Sivonen K, 1996. Occurrence of novel cyclic peptides together with microcystins from toxic cyanobacteria, *Anabaena* species. In: Yasumoto, T., Oshima, Y., Fukuyo, Y. (Eds.), Harmful and Toxic Algal Blooms. Intergovermental Oceanographic Commission of UNESCO, Paris, pp. 559-562.

[0341] Gerry, N. P. et al. Universal DNA mnicroarray method for multiplex detection of low abundance point mutations. J. Mol. Biol. 292, 251-262 (1999).

[0342] Gilroy, D. J., Kauffman, K. W., Hall, R. A., Huang, X. & Chu, F. S. Assessing potential health risks from microcystin toxins in blue-green algae dietary supplements. *Environ. Health Perspect.* 108,435439 (2000).

[0343] Giovannoni, S. J., E. F. DeLong, T. M. Schmidt, and N. R. Pace. 1990. Tangential flow filtration and preliminary phylogenetic analysis of marine picoplankton. Appl. Environ. Microbiol. 56:2572-2575.

[0344] Goldberg, J., Huang, H. B., Kwon, Y. G., Greengard, P., Nairn, A-C., Kuriyan, J., 1995. Three-dimensional structure of the catalytic subunit of protein serine/threonine phosphatase-1. Nature 376, 745-53.

[0345] Grüntzig, V., S. C. Nold, J. Zhou, and J. M. Tiedje. 2001. *Pseudomonas stutzeri* nitrite reductase gene abundance in environmental samples measured by real-time PCR. Appl. Environ. Microbiol. 67:760-768.

[0346] Gugger, M., C. Lyra, P. Henriksen, A. Couté, J.-F. Humbert, and K Sivonen. 2002. Phylogenetic comparison of the cyanobacterial genera *Anabaena* and *Aphanizomenon*. Int. J. Syst. Evol. Microbiol. 52:1867-1880.

[0347] Hall, T. A. BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. *Nucl. Acids. Symp. Ser.* 41, 95-98 (1999).

[0348] Harada, K.-I., F. Kondo, and L. Lawton. 1999. Laboratory analysis of cyanotoxins. p. 369-405. In I. Chorus and J. Bartram (ed.), Toxic cyanobacteria in water. E & FN Spon. London and New York.

[0349] Heckman, D. S. et al. Molecular evidence for the early colonization of land by fungi and plants. *Science* 293, 1129-1133 (2001).

[0350] Heid, C. A., J. Stevens, K J. Livak, and P. M. Williams. 1996. Real time quantitative PCR. Genome Res. 6:986-994.

[0351] Herdman, M., M. Janvier, R. Rippka, and R. Y. Stanier. 1979. Genome size of cyanobacteria. J. Gen. Microbiol. 111:73-85.

[0352] Hisbergues, M., Christiansen, G., Rouhiainen, L., Sivonen, K & Börner, T. PCR-based identification of microcystin producing genotypes of different cyanobacterial genera: application to environmental samples. *Arch Microbiol.* 180, 402-410 (2003).

[0353] Hopwood, D. A. Genetic contributions to understanding polyketide synthases. *Chem. Rev.* 97, 2465-2497 (1997).

[0354] Ikeda, H., Nonomiya, T., Usami, M., Ohta, T., Omura, S., 1999. Organization of the biosynthetic gene cluster for the polyketide anthelmintic macrolide avermectin in *Streptomyces avermitilis*. Proc. Natl. Acad. Sci. USA 96, 9509-9514.

[0355] Kaebernick, M., T. Rohrlack, K Chiistoffersen, and B. A. Neilan. 2001. A spontaneous mutant of microcystin biosynthesis: genetic characterization and effect on *Daphnia*. Environ. Microbiol. 3:669-679.

[0356] Kagan, R. M., Clarke, S., 1994. Widespread occurrence of three sequence motifs in diverse S-adenosylmethionine-dependent methyltransferases suggests a common structure for these enzymes. Arch. Biochem. Biophys. 310, 417-427.

[0357] Kohli, R. M., Trauger, J. W., Schwarzer, D., Marahiel, M. A., Walsh, C. T., 2001. Generality of peptide cyclization catalyzed by isolated thioesterase domains of nonribosomal peptide. Biochemistry 40, 7099-7108.

[0358] Kotai, J., 1972. Instructions for preparation of modified nutrient solution Z8 for algae, publication B-11/69, Norwegian Institute for Water Research. Blindern, Oslo.

[0359] Kuiper-Goodman, T., Falconer, I. & Fitzgerald, J. in *Toxic Cyanobacteria in Water. A Guide to their Public Health Consequences, Monitoring and Management* (eds Chorus, I. & Bartham, J.) 113-153 (E and FN Spoon, London, 1999).

[0360] Kurmayer, R., E. Dittmann, J. Fastner, and I. Chorus. 2002. Diversity of microcystin genes within a population of the toxic cyanobacterium *Microcystis* spp. in Lake Wannsee (Berlin, Germany). Microb. Ecol. 43:107-118.

[0361] Labarre, J., F. Chauvat, and P. Thuriaux. 1989; Insertional mutagenesis by random cloning of antibiotic resistance genes into the genome of the cyanobacterium *Synechocystis* strain PCC 6803. J. Bacteriol. 171:3449-3457.

[0362] Lahti, K, J. Rapala, M. Färdig, M. Niemelä, and K Sivonen. 1997. Persistence of cyanobacterial hepatotoxin, microcystin-LR in particulate material and dissolved in lake water. Wat. Res. 31:1005-1012.

[0363] Lee, S. J., M.-H. Jang, H.-S. Kim, B.-D. Yoon, and H.-M. Oh. 2000. Variation of microcystin content of *Microcystis aeruginosa* relative to medium N:P ratio and growth stage. J. Appl. Microbiol. 89:323-329.

[0364] Lepere, C., Wilmotte, A. & Meyer B. Molecular diversity of *Microcystis* strains (Cyanophyceae, Chroococcales) based on 16S rDNA sequences. *Syst. Geogr. Pl.* 70, 275-283 (2000).

[0365] Lindroos K et al, 2002 Nucleic Acid Research vol 30, no 14e70.

[0366] Lyra, C., Suomalainen, S., Gugger, M., Vezie, C., Sundman, P., Paulin, L., Sivonen, K, 2001. Molecular characterization of planktic cyanobacteria of *Anabaena, Aphanizomenon, Microcystis* and *Planktothrix* genera Int. J. System. Evol. Microbiol. 51, 513-526.

[0367] MacKintosh, C., Beattie, K. A., Klumpp, S., Cohen, P., Codd, G. A., 1990. Cyanobacterial microcystin-LR is a potent and specific inhibitor of protein phosphatase 1 and 2A from both mammals and higher plants. FEBS Lett. 264, 187-192;

[0368] Marahiel, M. A., Stachelhaus, T., Mootz, H. D., 1997. Modular peptide synthetases involved in nonribosomal peptide synthesis. Chem. Rev. 97, 2651-2673.

[0369] Mehta, P. K, Hale, T. I., Christen, P., 1993. Aminotranferases: demonstration of homology and division into evolutionary subgroups. Eur. J. Biochem. 214, 549-561.

[0370] Moffitt, M. C. & Neilan, B. A. On the presence of peptide synthetase and polyketide synthase genes in the cyanobacterial genus *Nodularia. FEMS Microbiol. Lett.* 196, 207-214 (2001).

[0371] Moon-van der Staay, S-Y., Wachter, R. & Vaulot, D. Oceanic 18S rDNA sequences from picoplankton reveal unsuspected eukaryotic diversity. *Nature* 409, 607-610 (2001).

[0372] Moore, R. E., Chen, J. L, Moore, B. S., Patterson, G. M. L., Charmichael, W. W., 1991. Biosynthesis of microcystin-LR. Origin of the carbons in the Adda and Masp units. J. Am Chem. Soc. 113, 5083-5084.

[0373] Namikoshi, M., K L. Rinehart, R. Sakai, R. R. Stotts, A. M. Dahlem, V. R. Beasley, W. W. Carmichael, and W. R. Evans. 1992. Identification of 12 hepatotoxins from a Horner Lake bloom of the cyanobacteria *Microcystis aeruginosa, Microcystis viridis*, and *Microcystis wesenbergii*: nine new microcystins. J. Org. Chem. 57:866-872.

[0374] Namikoshi, M., Rinehart, KL., 1996. Bioactive combounds produced by cyanobacteria. J. Ind. Microbiol. 17, 373-384.

[0375] Neilan, B. A, D. Jacobs, T. del Dot, L. L. Blackall, P. R. Hawkins, P. T. Cox, and A. E. Goodman. 1997. rRNA sequences and evolutionary relationships among toxic and nontoxic cyanobacteria of the genus *Microcystis*. hit. J. Syst. Bacteriol. 47:693-697.

[0376] Neilan, B. A. et al. Nonribosomal peptide synthesis and toxigenity of cyanobacteria. *J. Bacteriol.* 181, 4089-4097 (1999).

[0377] Nishizawa, T., M. Asayama, K. Fujii, K–1. Harada, and M. Shirai. 1999. Genetic analysis of the peptide synthetase genes for a cyclic heptapeptide microcystin in *Microcystis* spp. J. Biochem. 126:520-529.

[0378] Nishizawa, T., Ueda, A., Asayama, M., Fujii, K., Harada, K-, Ochi, K, Shirai, M., 2000. Polyketide synthase gene coupled to the peptide synthetase module involved in the biosynthesis of the cyclic heptapeptide microcystin. J. Biochem. 127, 779-789.

[0379] Nishizawa, T., Asayama, M., Shirai, M., 2001. Cyclic heptapeptide microcystin biosynthesis requires the glutamate racemase gene. Microbiology 147, 1235-1241.

[0380] Normeman, D., and P. V. Zimba. 2002. A PCR-based test to assess the potential for microcystin occurrence in channel catfish production ponds. J. Phycol. 38:230-233.

[0381] Ohtake, A., M. Shirai, T. Aida, N. Mori, K.-I. Harada, KI Matsuura, M. Suzuki, and M. Nakano. 1989. Toxicity of *Microcystis* species isolated from natural blooms and purification of the toxin. Appl. Environ. Microbiol. 55:3202-3207.

[0382] Orr, P. T., and G. J. Jones. 1998. Relationship between microcystin production and cell division rates in nitrogen-limited *Microcystis* aeruginosa cultures. Limnol. Oceanogr. 43:1604-1614.

[0383] Otsuka, S. et al. Phylogenetic relationships between toxic and non-toxic strains of the genus *Microcystis* based on 16S to 23S internal transcribed spacer sequence. *FEMS Microbiol. Lett.* 172, 15-21 (1999).

[0384] Pan, H., L. Song, Y. Liu, and T. Börner. 2002. Detection of hepatotoxic *Microcystis* strains by PCR with intact cells from both culture and environmental samples. Arch. Microbiol. 178:421-427.

[0385] Rapala, J., K Sivonen, C. Lyra, and S. L Niemeli. 1997. Variation of microcystins, cyanobacterial hepatotoxins in *Anabaena* spp. as a function of growth stimuli. Appl. Environ. Microbiol. 63:2206-2212.

[0386] Rapala, J., and K Sivonen. 1998. Assessment of environmental conditions that favor hepatotoxic and neurotoxic *Anabaena* spp. strains cultured under light limitation at different temperatures. Microb. Ecol. 36:181-192.

[0387] Repka, S., J. Mehtonen, J. Vaitomaa, L. Saari, and K. Sivonen. 2001. Effects of nutrients on growth and nodularin production of *Nodularia* strain GR8b. Microb. Ecol. 42:606-613.

[0388] Ressom, R. et al. *Health Effects of Toxic Cyanobacteria (Blue-green Algae)* (Australian Government Publishing Service, Canberra, 1994).

[0389] Reynolds, C. S. 1997. Vegetation processes in the pelagic: a model for ecosystem theory. Excellence in Ecology, Ecology Institute, Oldendorf/Luhe, 371 p.

[0390] Rippka, R., and M. Herdman. 1992. Pasteur culture collection of cyanobacterial strains in axenic culture. Catalogue & taxonomic handbook, volume I: catalogue of strains 1992/1993. Paris, Institut Pasteur. p. 103.

[0391] Rippka, R., Castenholz, R. W., Iteman, I., Herdman, M., 2001. Form-genus I. *Anabaena* Bory de St. Vincent 1822 sensu Rippka, Demelles, Waterbury, Herdman and Stanier 1979. In: Boone, D. R., Castenholz, R. W. (Eds.), Bergey's Manual of Systematic Bacteriology. Springer-Verlag, New York, pp. 566-568. 2nd edn, vol. 1.

[0392] Rouhiainen, L., Sivonen, K, Buikema, W. J., Haselkorn, R., 1995. Characterization of toxin-producing cyanobacteria by using an oligonucleotide probe containing a tandemly repeated heptamer. J. Bacteriol. 177, 6021-6026.

[0393] Rouhiainen, L., Paulin, L., Suomalainen S., Hyytiäinen, H., Buikema, W., Haselkorn, R., Sivonen, K, 2000. Genes encoding synthetases of cyclic depsipeptides, anabaenopeptilides, in *Anabaena* strain 90. Mol. Microbiol. 37, 156-167.

[0394] Saitou N., Nei M. The neighbor-joining method: a new method for reconstructing phylogenetic trees. Mol Biol Evol. 4, 406-25 (1987).

[0395] Sambrook, J., Fritsch, E. F., Maniatis, T., 1989. Molecular cloning: A Laboratory Manual. Cold Spring Harbor Laboratory Press, Plainview, N.Y., 2nd edn.

[0396] Scrutton, N. S., Beny, A., Perham, R. N., 1990. Redesign of the coenzyme specificity of a dehydrogenase by protein engineering. Nature 343, 38-43.

[0397] Silakowski, B., Nordsiek, G., Kunze, B., Blocker, H., Müller R., 2001. Novel features in a combined polyketide synthase/non-ribosomal peptide synthetase: the myxalamid biosynthetic gene cluster of the myxobacterium *Stigmatella aurantiaca* Sga 15. Chem. Biol. 8, 59-69.

[0398] Sivonen, K., K Himberg, R. Luukkainen, S. L. Niemelä, G. K Poon, and G. A. Codd. 1989. Preliminary characterization of neurotoxic cyanobacteria blooms and strains from Finland Toxic. Assess. 4:339-352.

[0399] Sivonen, K, W. W. Carmichael M. Namikoshi, K. L. Rinehart, A. M. Dahlem, and S. I. Niemelä. 1990. Isolation and characterization of hepatotoxic microcystin homologs from the filamentous freshwater cyanobacterium *Nostoc* sp. strain 152. Appl. Environ. Microbiol. 56:2650-2657.

[0400] Sivonen, K, Namikoshi, M., Evans, W. R., Carmichael W. W., Sun, F., Rouhiainen, L., Luukkainen, R., Rinehart, K. L., 1992. Isolation and characterization of a variety of microcystins from seven strains of the cyanobacterial genus *Anabaena*. Appl. Environ. Microbiol. 58, 2495-2500.

[0401] Sivonen, K, M. Namikoshi, R. Luukkainen, M. Fardig, L. Rouhiainen, W. R. Evans, W. W. Carmichael, K L. Rinehart, and S. I. Niemela. 1995. Variation of cyanobacterial hepatotoxins in Finland, p. 163-169. In M. Munavar and M. Luotola (ed.), The contaminants in the nordig ecosystem: dynamics, processes & fate. Ecovision world monograph series, Asssterdam, the Neatherlands.

[0402] Sivonen, K, Jones, G., 1999. Cyanobacterial toxins. In: Chorus, I., Bertram J. (Eds.), Toxic Cyanobacteria in Water: A Guide to their Public Health Consequences, Monitoring and Management. E. & F. N. Spon, London, pp. 41-111.

[0403] Stachelhaus, T., Mootz, H. D., Marahiel M. A., 1999. The specificity-conferring code of adenylation domains in normbosomal peptide synthetases. Chem. Biol. 6, 493-505.

[0404] Suzuki, M. T., and S. J. Giovannoni. 1996. Bias caused by template annealing in the amplification of mixtures of 16S rRNA genes by PCR. Appl. Environ. Microbiol. 62:625-630.

[0405] Swofford, D. L., Olsen, G. J., Waddell, P. J. & Hillis, D. M. in *Molecular Systematics* (eds Hillis, D. M., Moritz, C. & Mable, B. K) 407-514 (Sinauer, Sunderland, Mass., 1996).

[0406] Swofford, D. L. PAUP*. *Phylogenetic Analysis Using Parsimony* (**and Other Methods*) Version 4.0b8 (Sinaeur, Sunderland, Mass., 2001).

[0407] Tang, L., Yoon, Y. J., Choi, C-Y., Hutchinson, C. R., 1998. Characterization of the enzymatic domains in the modular polyketide synthase involved in rifamycin B biosynthesis by *Amycolatopsis mediterranei*. Gene 216, 255-265.

[0408] Taton, A., S. Grubisic, E. Brambilla, R. De Wit and A. Wilmotte (2003). Cyanobacterial Diversity in Natural and Artificial Microbial Mats of Lake Fryxell (McMurdo Dry Valleys, Antarctica): a Morphological and Molecular Approach. Appl. Environ. Microbiol. 69, 5157-5169.

[0409] Thompson J D, Higgins D G, Gibson T J. CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. Nucleic Acids Res 22, 4673-80 (1994).

[0410] Tillett, D., Dittmann, E., Erhard, M., von Döhren, H., Borner, T., Neilan, B. A., 2000. Structural organization of microcystin biosynthesis in *Microcystis aeruginosa* PCC7806: an integrated peptide-polyketide synthetase system. Chem. Biol. 7, 753-764.

[0411] Tillett, D., Parker, D. L. & Neilan, B. A. Detection of toxigenicity by a probe for the microcystin synthetase A gene (mcyA) of the cyanobacterial genus *Microcystis*: comparison of toxicities with 16S rRNA and phycocyanin operon (phycocyanin intergenic spacer) phylogenies. *Appl. Environ.* Microbiol. 67, 2810-2818 (2001).

[0412] Tsuge, K, Akiyama, T., Shoda, M., 2001. Cloning, sequencing, and characterization of the iturin A operon. J. Bacteriol. 183, 6265-6273.

[0413] Utermöhl, H. 1958. Zur Vervollkommnung der quantitativen phytoplankton-methodik. Mitt. Int. Ver. Limnol. 9:1-38.

[0414] Utkilen, H. & Gjøhme, N. Iron-stimulated toxin production in *Microcystis aeruginosa. Appl. Environ. Microbiol.* 61, 797-800 (1995).

[0415] Vasconcelos, V. M., K Sivonen, W. R. Evans, W. W. Carmichael, and M. Namikoshi. 1995. Isolation and characterization of microcystins (haptapeptide hepatotoxins) from Portuguese strains of *Microcystis aeruginosa* KUTZ. emend ELEKIN. Arch. Hydrobiol. 134:295-305.

[0416] Vezie, C., L. Brient, K. Sivonen, G. Bertru, J.-C. Lefeuvre, and M. Salkinoja-Salonen. 1998. Variation of microcystin content of cyanobacterial blooms and isolated strains in Lake Grand-Lieu (France). Microb. Ecol. 35:126-135.

[0417] Vézie, C., J. Rapala, J. Vaitomaa, J. Seitsonen, and K. Sivonen. 2002. Effect of nitrogen and phosphorus on growth of toxic and nontoxic *Microcystis* strains and on intracellular microcystin concentrations. Microb. Ecol. 43:443-454.

[0418] Wawrik, B., J. H. Paul, and F. R. Tabita. 2002. Real-time PCR quantification of rbcL (ribulose-1,5-bisphosphate carboxylase/oxygenase) mRNA in diatoms and pelagophytes. Appl. Environ. Microbiol. 68:3771-3779.

[0419] Willén, T. 1962. Studies on the phytoplankton of some lakes connected with or recently isolated from the Baltic. Oikos 13:169-199.

[0420] Wintzingerode F., U. B. Gobel, and E. Stackebrandt. 1997. Determination of microbial diversity in environmental samples: pitfalls of PCR-based rRNA analysis. FEMS Microbiol. Rev. 21:213-229.

SEQUENCE LISTING

<160> NUMBER OF SEQ ID NOS: 170

<210> SEQ ID NO 1
<211> LENGTH: 769
<212> TYPE: DNA
<213> ORGANISM: Anabaena Ana90

<400> SEQUENCE: 1

agcattagca gttggttatc gtaatcttcc tgaaattacc agagaaaaat ttcaacccaa 60

ctttttaaat tcagaaaaaa ttctctttag aacgggagat ttaggtaaac aaattgctcc 120

aggtgtgatt gaatttatag ggcgaaaaga taatcaagtt aaggtcaatg gctatcgtgt 180

agatccagga gaaattgaat accaaattag ccgctatgcc gagattgaga aagcaattgt 240

cttacctata gaggtaaata accaaattca attatctgct tattgtcaaa ctgataaaga 300

tataaaattt tctgaaatca gagaattttt agctaaatat ttgccagttt acatgattcc 360

tagttccttt atcttcttaa agcaatttcc cttaactaaa catggcaaat ttgacttgcg 420

atcgctcatc gctctcaagc caacagatca attaacacaa gtctcttata ctgcaccgcg 480

taatacttta gaatcaaagc tagtccatat ttgggaaaaa attctcacta aacatcccat 540

tggaattttt gataactttt ttgaaatcgg cggacactct ctgctccttt ctagagtagt 600

cactcacgtc cataaagaat taaatgtatt agttaaatta gctgatttct tcaaagttcc 660

cacaattctt ggattagcag ctttaatatc taaagctcaa tctaactatc aagaacccat 720

accagcaata actcaacaag aatcttatcc catgtctcat ggacaacgc 769

<210> SEQ ID NO 2
<211> LENGTH: 769
<212> TYPE: DNA
<213> ORGANISM: Anabaena Ana 299B

<400> SEQUENCE: 2

agcattagca gttggttatc gtaatcttcc tgaaattacc agagaaaaat ttcaacccaa 60

ctttttaaat tcagaaaaaa ttctctttag aacgggagat ttaggtaaac aaattgctcc 120

aggtgtgatt gaatttatag ggcgaaaaga taatcaagtt aaggtcaatg gctatcgtgt 180

agatccagga gaaattgaat accaaattag ccgctatgcc gagattgaga aagcaattgt 240

cttacctata gaggtaaata accaaattca attatctgct tattgtcaaa ctgataaaga 300

tataaaattt tctgaaatca gagaattttt agctaaatat ttgccagttt acatgattcc 360

tagttccttt atcttcttaa agcaatttcc cttaactaaa catggcaaat ttgacttgcg 420

atcgctcatc gctctcaagc caacagatca attaacacaa gtctcttata ctgcaccgcg 480

taatacttta gaatcaaagc tagtccatat ttgggaaaaa attctcacta aacatcccat 540

tggaattttt gataactttt ttgaaatcgg cggacactct ctgctccttt ctagagtagt 600

cactcacgtc cataaagaat taaatgtatt agttaaatta gctgatttct tcaaagttcc 660

cacaattctt ggattagcag ctttaatatc taaagctcaa tctaactatc aagaacccat 720

accagcaata actcaacaag aatcttatcc catgtctcat ggacaacgc 769

<210> SEQ ID NO 3
<211> LENGTH: 769
<212> TYPE: DNA

-continued

<213> ORGANISM: Anabaena Ana315

<400> SEQUENCE: 3

agcattagca gttggttatc gtaatcttcc tgaaattacc agagaaaaat ttcaacccaa 60

ctttttaaat tcagaaaaaa ttctctttag aacgggagat ttaggtaaac aaattgctcc 120

aggtgtgatt gaatttatag ggcgaaaaga taatcaagtt aaggtcaatg gctatcgtgt 180

agatccagga gaaattgaat accaaattag ccgctatgcc gagattgaga aagcaattgt 240

cttacctata gaggtaaata accaaattca attatctgct tattgtcaaa ctgataaaga 300

tataaaattt tctgaaatca gagaattttt agctaaatat ttgccagttt acatgattcc 360

tagttccttt atcttcttaa agcaatttcc cttaactaaa catggcaaat ttgacttccg 420

atcgctcatc gctctcaagc caacagatca attaacacaa gtctcttata ctgcaccgcg 480

taatacttta gaatcaaagc tagtccatat ttgggaaaaa attctcacta aacatcccat 540

tggaattttt gataactttt ttgaaatcgg cggacactct ctgctccttt ctagagtagt 600

cactcacgtc cataaagaat taaatgtatt agttaaatta gctgatttct tcaaagttcc 660

cacaattctt ggattagcag ctttaatatc taaagctcaa tctaactatc aagaacccat 720

accagcaata actcaacaag aatcttatcc catgtctcat gggcaacgc 769

<210> SEQ ID NO 4
<211> LENGTH: 769
<212> TYPE: DNA
<213> ORGANISM: Anabaena Ana PH256

<400> SEQUENCE: 4

agcattagca gttggttatc gtaatcttcc tgaaattacc agagaaaaat ttcaacccaa 60

ctttttaaat tcagaaaaaa ttctctttag aacgggagat ttaggtaaac aaattgctcc 120

aggtgtgatt gaatttatag ggcgaaaaga taatcaagtt aaggtcaatg gctatcgtgt 180

agatccagga gaaattgaat accaaattag ccgctatgcc gagattgaga aagcaattgt 240

cttacctata gaggtaaata accaaattca attatctgct tattgtcaaa ctgataaaga 300

tataaaattt tctgaaatca gagaattttt agctaaatat ttgccagttt acatgattcc 360

tagttccttt atcttcttaa agcaatttcc cttaactaaa catggcaaat ttgacttccg 420

atcgctcatc gctctcaagc caacagatca attaacacaa gtctcttata ctgcaccgcg 480

taatacttta gaatcaaagc tagtccatat ttgggaaaaa attctcacta aacatcccat 540

tggaattttt gataactttt ttgaaatcgg cggacactct ctgctccttt ctagagtagt 600

cactcacgtc cataaagaat taaatgtatt agttaaatta gctgatttct tcaaagttcc 660

cacaattctt ggattagcag ctttaatatc taaagctcaa tctaactatc aagaacccat 720

accagcaata actcaacaag aatcttatcc catgtctcat ggacaacgc 769

<210> SEQ ID NO 5
<211> LENGTH: 769
<212> TYPE: DNA
<213> ORGANISM: Anabaena Ana 318

<400> SEQUENCE: 5

agcattagca gttggttatc gtaatcttcc tgaaattacc agagaaaaat ttcaacccaa 60

ctttttaaat tcagaaaaaa ttctctttag aacgggagat ttaggtaaac aaattgctcc 120

-continued

aggtgtgatt gaatttatag ggcgaaaaga taatcaagtt aaggtcaatg gctatcgtgt 180

agatccagga gaaattgaat accaaattag ccgctatgcc gagattgaga aagcaattgt 240

cttacctata gaggtaaata accaaattca attatctgct tattgtcaaa ctgataaaga 300

tataaaattt tctgaaatca gagaattttt agctaaatat ttgccagttt acatgattcc 360

tagttccttt atcttcttaa agcaatttcc cttaactaaa catggcaaat ttgacttgcg 420

atcgctcatc gctctcaagc caacagatca attaacacaa gtctcttata ctgcaccgcg 480

taatacttta gaatcaaagc tagtccatat ttgggaaaaa attctcacta aacatcccat 540

tggaattttt gataactttt ttgaaatcgg cggacactct ctgctccttt ctagagtagt 600

cactcacgtc cataaagaat taaatgtatt agttaaatta gctgatttct tcaaagttcc 660

cacaattctt ggattagcag ctttaatatc taaagctcaa tctaactatc aagaacccat 720

accagcgata actccacaag catcttatcc catgtctcat ggacaacgc 769

<210> SEQ ID NO 6
<211> LENGTH: 769
<212> TYPE: DNA
<213> ORGANISM: Anabaena Ana202A1

<400> SEQUENCE: 6

agcattagca gttggttatc gtaatcttcc tgaaattacc agagaaaaat ttcaacccaa 60

ctttttaaat tcagaaaaaa ttctctttag aacgggagat ttaggtaaac aaattgctcc 120

aggtgtgatt gaatttatag ggcgaaaaga taatcaagtt aaggtcaatg gctatcgtgt 180

agatccagga gaacttgaat accaaattag ccgctatgcc gagattgaga aagcaattgt 240

cttacctata gaggtaaata accaaattca attatctgct tattgtcaaa ctgataaaga 300

tataaaaatt tctgaaatca gagaattttt agctaaatat tcgccagttt acatgattcc 360

tagttccttt atcttcttaa agcaatttcc cttaactaaa catggcaaac ttgacttgcg 420

atcgctcgtc gctctcaagc cgacagatca actaacacaa gtctcttata ctgcaccgcg 480

taatacttta gaatcaaagc tagtccatat ttgggaaaaa attctcacta aacatcccat 540

tggaattttt gataactttt ttgaaatcgg cggacactct ctgctccttt ctagagtagt 600

cactcacgtc cataaagaat taaatgtatt agttaaatta gctgatttct tcaaagttcc 660

cacaattctt ggattagcag ctttaatatc taaagctcaa tctaactatc aagaacccat 720

accagcaata actcaacaag aatcttatcc catgtctcat ggacaacgc 769

<210> SEQ ID NO 7
<211> LENGTH: 769
<212> TYPE: DNA
<213> ORGANISM: Anabaena Ana202A2-41

<400> SEQUENCE: 7

agcattagca gttggttatc gtaatcttcc tgaaattacc agagaaaaat ttcaacccaa 60

ctttttaaat tcagaaaaaa ttctctttag aacgggagat ttaggtaaac aaattgctcc 120

aggtgtgatt gaatttatag ggcgaaaaga taatcaagtt aaggtcaatg gctatcgtgt 180

agatccagga gaacttgaat accaaattag ccgctatgcc gagattgaga aagcaattgt 240

cttacctata gaggtaaata accaaattca attatctgct tattgtcaaa ctgataaaga 300

tataaaaatt tctgaaatca gagaattttt agctaaatat tcgccagttt acatgattcc 360

-continued

tagttccttt atcttcttaa agcaatttcc cttaactaaa catggcaaac ttgacttgcg 420

atcgctcgtc gctctcaagc cgacagatca actaacacaa gtctcttata ctgcaccgcg 480

taatacttta gaatcaaagc tagtccatat ttgggaaaaa attctcacta aacatcccat 540

tggaattttt gataactttt ttgaaatcgg cggacactct ctgctccttt ctagagtagt 600

cactcacgtc cataaagaat taaatgtatt agttaaatta gctgatttct tcaaagttcc 660

cacaattctt ggattagcag ctttaatatc taaagctcaa tctaactatc aagaacccat 720

accagcaata actcaacaag aatcttatcc catgtctcat ggacaacgc 769

<210> SEQ ID NO 8
<211> LENGTH: 769
<212> TYPE: DNA
<213> ORGANISM: Anabaena Ana66A

<400> SEQUENCE: 8

agcattagca gttggttatc gtaatcttcc tgaaattacc agagaaaaat ttcaacccaa 60

ctttttaaat tcaggaaaaa ttctctttag aacgggagat ttaggtaaac aaattgctcc 120

aggtgtgatt gaatttatag ggcgaaaaga taatcaagtt aaggtcaatg gctatcgtgt 180

agatccagga gaacttgaat accaaattag ccgctatgcc gagattgaga aagcaattgt 240

cttacctata gaggtaaata accaaattca attatctgct tattgtcaaa ctgataaaga 300

tataaaaatt tctgaaatca gagaattttt agctaaatat tcgccagttt acatgattcc 360

tagttccttt atcttcttaa agcaatttcc cttaactaaa catggcaaac ttgacttgcg 420

atcgctcgtc gctctcaagc cgacagatca actaacacaa gtctcttata ctgcaccgcg 480

taatacttta gaatcaaagc tagtccatat ttgggaaaaa attctcacta aacatcccat 540

tggaattttt gataactttt ttgaaatcgg cggacactct ctgctccttt ctagagtagt 600

cactcacgtc cataaagaat taaatgtatt agttaaatta gctgatttct tcaaagttcc 660

cacaattgtt ggattagcag ctttaatatc taaagctcaa tctaactatc aagaacccat 720

accagcgata gctccacaag catcttatcc catgtctcat gggcaacgc 769

<210> SEQ ID NO 9
<211> LENGTH: 769
<212> TYPE: DNA
<213> ORGANISM: Anabaena Ana83-1

<400> SEQUENCE: 9

agcattagca gttggttatc gtaatcttcc tgaaattacc agagaaaaat ttcaacccaa 60

ctttttaaat tcagaaaaaa ttctctttag aacgggagat ttaggtaaac aaattgctcc 120

aggtgtgatt gaatttatag ggcgaaaaga taatcaagtt aaggtcaatg gctatcgtgt 180

agatccagga gaacttgaat accaaattag ccgctatgcc gagattgaga aagcaattgt 240

cttacctata gaggtaaata accaaattca attatctgct tattgtcaaa ctgataaaga 300

tataaaaatt tctgaaatca gagaattttt agctaaatat tcgccagttt acatgattcc 360

tagttccttt atcttcttaa agcaatttcc cttaactaaa catggcaaac ttgacttgcg 420

atcgctcgtc gctctcaagc cgacagatca actaacacaa gtctcttata ctgcaccgcg 480

taatacttta gaatcaaagc tagtccatat ttgggaaaaa attctcacta aacatcccat 540

tggaattttt gataactttt ttgaaatcgg cggacactct ctgctccttt ctagagtagt 600

-continued

cactcacgtc cataaagaat taaatgtatt agttaaatta gctgatttct tcaaagttcc 660

cacaattctt ggattagcag ctttaatatc taaagctcaa tctaactatc aagaacccat 720

accagcaata actcaacaag aatcttatcc catgtctcat ggacaacgc 769

<210> SEQ ID NO 10
<211> LENGTH: 769
<212> TYPE: DNA
<213> ORGANISM: Anabaena consensus

<400> SEQUENCE: 10

agcattagca gttggttatc gtaatcttcc tgaaattacc agagaaaaat ttcaacccaa 60

ctttttaaat tcagaaaaaa ttctctttag aacgggagat ttaggtaaac aaattgctcc 120

aggtgtgatt gaatttatag ggcgaaaaga taatcaagtt aaggtcaatg gctatcgtgt 180

agatccagga gaamttgaat accaaattag ccgctatgcc gagattgaga aagcaattgt 240

cttacctata gaggtaaata accaaattca attatctgct tattgtcaaa ctgataaaga 300

tataaaattt tctgaaatca gagaattttt agctaaatat ttgccagttt acatgattcc 360

tagttccttt atcttcttaa agcaatttcc cttaactaaa catggcaaat ttgacttgcg 420

atcgctcatc gctctcaagc cracagatca aytaacacaa gtctcttata ctgcaccgcg 480

taatacttta gaatcaaagc tagtccatat ttgggaaaaa attctcacta aacatcccat 540

tggaattttt gataactttt ttgaaatcgg cggacactct ctgctccttt ctagagtagt 600

cactcacgtc cataaagaat taaatgtatt agttaaatta gctgatttct tcaaagttcc 660

cacaattctt ggattagcag ctttaatatc taaagctcaa tctaactatc aagaacccat 720

accagcaata actcaacaag aatcttatcc catgtctcat ggacaacgc 769

<210> SEQ ID NO 11
<211> LENGTH: 772
<212> TYPE: DNA
<213> ORGANISM: Microcystis Mpcc7806-Tillet

<400> SEQUENCE: 11

cgccctagca tcgggttatc ataaccaacc cgaaatgact caagaaaaat ttaaacctag 60

ctttcttgat gagacaaaaa ctctctttag aaccggcgat ttaggcaagc aaactgctcc 120

cggtatcatt gagtttatgg gacgaaaaga taatcaagtt aaggtcaatg gttatcgaat 180

tgaccccgga gaaattgaat atcaattgac tcgttatgct cccattgaaa gagcgattgt 240

tttacccgtt caagttaata atcaaactca attatctgct tactgtcaaa cagacaaaac 300

tctagaaatt gctgagattc gagaattact tgccaaattt ttaccagttt atatgattcc 360

gagttacttt attttttaa agcaattccc cttaactcga catgggaaac ttgacctgca 420

ctccctgaga gaactcagag aaactggtaa atctctggtg aattctaatt atgttgcacc 480

ccggaattat ttagaatcca atctcgttag tatctgggaa aaaattctct ctaaacatcc 540

tatcggtatt tttgataact tctttgaaat tggcggtcat tctctactct tatcaagggt 600

tgtaacccgg gttcataaag aactaaatgt atccgtaaaa ttggctgact tctttaaagt 660

tccaaccatt gctggattgg ctactttaat ctcccagact caatacaatt atcaagaacc 720

catttcggca attccccccc aaaaatccta tccgatgtct catggtcagc gt 772

<210> SEQ ID NO 12

-continued

<211> LENGTH: 772
<212> TYPE: DNA
<213> ORGANISM: Microcystis M205

<400> SEQUENCE: 12

cgccctagca tcgggttatc ataaccaacc cgaaatgact caagaaaaat ttaaacctag 60

ctttcttgat gagacaaaaa ctctctttag aaccggcgat ttaggcaaac aaactgctcc 120

aggtatcatt gagtttatgg gacgaaaaga taatcaagtt aaggtcaatg gttatcgaat 180

tgaccccgga gaaattgaat atcaattgac tcgttatgct cccattgaaa gagcgattgt 240

tttacccgtt caagttaata atcaaactca attatctgct tactgtcaaa cagacaaaac 300

tctagaaatt gctgagattc gagaattact tgccaaattt ttaccagttt atatgattcc 360

gagttacttt attttttaa agcaattccc cttaactcga catggaaaac ttgacctgca 420

ctccctgaga gaactcagag aaactggtaa atctctggtg aattctaatt acgttgcacc 480

ccggaattat ttagaatcca atctcgttag tatctgggaa aaaattctct ctaaacatcc 540

tatcggtatt tttgataact tctttgaaat tggcggtcat tctctactct tatcaagggt 600

tgtaacccgg gttcataaag aactaaatgt atccgtaaaa ttagctgact tctttaaagt 660

tccaaccgtt gctggattgg cgactttaat ctcccagact caatacaatt atcaagaacc 720

catttcggca attcccccc aaaaatctta tccgatgtct catggtcagc gt 772

<210> SEQ ID NO 13
<211> LENGTH: 772
<212> TYPE: DNA
<213> ORGANISM: Microcystis M98

<400> SEQUENCE: 13

cgccctagca tcgggttatc ataaccaacc cgaaatgact caagaaaaat ttaaacctag 60

ctttcttgat gagacaaaaa ctctctttag aaccggcgat ttaggcaagc aaactgctcc 120

gggtatcatt gagtttatgg gacgaaaaga taatcaagtt aaggtcaatg gttatcgaat 180

tgaccccgga gaaattgaat atcaattgac tcgttatgct cccattgaaa gagcgattgt 240

tttacccgtt caagttaata atcaaactca attatctgct tactgtcaaa cagacaaaac 300

tctagaaatt gctgagattc gagaattact tgccaaattt ttaccagttt atatgattcc 360

gagttacttt attttttaa agcaattccc cttaactcga catggaaaac ttgacctgca 420

ctccctgaga gaactcagag aaactggtaa atctctggtg aattctaatt acgttgcacc 480

ccggaattat ttagaatcca atctcgttag tatctgggaa aaaattctct ctaaacatcc 540

tatcggtatt tttgataact tctttgaaat tggcggtcat tctctactct tatcaagggt 600

tgtaacccgg gttcataaag aactaaatgt atccgtaaaa ttagctgact tctttaaagt 660

tccaaccgtt gctggattgg cgactttaat ctcccagact caatacaatt atcaagaacc 720

catttcggca attcccccc aaaaatctta tccgatgtct catggtcagc gt 772

<210> SEQ ID NO 14
<211> LENGTH: 772
<212> TYPE: DNA
<213> ORGANISM: Microcystis Mgl260735

<400> SEQUENCE: 14

cgccctagca tcgggttatc ataaccaacc cgaaatgact caagaaaaat ttaaacctag 60

-continued

ctttcttgat gagacaaaaa ctctctttag aaccggcgat ttaggcaaac aaactgctcc 120

aggtatcatt gagtttatgg gacgaaaaga taatcaagtt aaggtcaatg gttatcgaat 180

tgaccccgga gaaattgaat atcaattgac tcgttatgct cccattgaaa gagcgattgt 240

tttacccgtt caagttaata atcaaactca attatctgct tactgtcaaa cagacaaaac 300

tctagaaatt gctgagattc gagaattact tgccaaattt ttaccagttt atatgattcc 360

gagttacttt attttttaa agcaattccc cttaactcga catggaaaac ttgacctgca 420

ctccctgaga gaactcagag aaactggtaa atctctggtg aattctaatt acgttgcacc 480

ccggaattat ttagaatcca atctcgttag tatctgggaa aaaattctct ctaaacatcc 540

tatcggtatt tttgataact tctttgaaat tggcggtcat tctctactct tatcaagggt 600

tgtaacccgg gttcataaag aactaaatgt atccgtaaaa ttagctgact tctttaaagt 660

tccaaccgtt gctggattgg cgactttaat ctcccagact caatacaatt atcaagaacc 720

catttcggca attccccccc aaaaatctta tccgatgtct catggtcagc gt 772

<210> SEQ ID NO 15
<211> LENGTH: 772
<212> TYPE: DNA
<213> ORGANISM: Microcystis Mgl280646

<400> SEQUENCE: 15

cgccctagca tcgggttatc ataaccaacc cgaaatgact caagaaaaat ttaaacctag 60

ctttcttgat gagacaaaaa ctctctttag aaccggcgat ttaggcaagc aaactgctcc 120

gggtatcatt gagtttatgg gacgaaaaga taatcaagtt aaggtcaatg gttatcgaat 180

tgaccccgga gaaattgaat atcaattgac tcgttatgct cccattgaaa gagcgattgt 240

tttacccgtt caagttaata atcaaactca attatctgct tactgtcaaa cagacaaaac 300

tctagaaatt gctgagattc gagaattact tgccaaattt ttaccagttt atatgattcc 360

gagttacttt attttttaa agcaattccc cttaactcga catggaaaac ttgacctgca 420

ctccctgaga gaactcagag aaactggtaa atctctggtg aattctaatt acgttgcacc 480

ccggaattat ttagaatcca atctcgttag tatctgggaa aaaattctct ctaaacatcc 540

tatcggtatt tttgataact tctttgaaat tggcggtcat tctctactct tatcaagggt 600

tgtaacccgg gttcataaag aactaaatgt atccgtaaaa ttagctgact tctttaaagt 660

tccaaccgtt gctggattgg cgactttaat ctcccagact caatacaatt atcaagaacc 720

catttcggca attccccccc aaaaatctta tccgatgtct catggtcagc gt 772

<210> SEQ ID NO 16
<211> LENGTH: 772
<212> TYPE: DNA
<213> ORGANISM: Microcystis Mpcc7941

<400> SEQUENCE: 16

cgccctagca tcgggttatc ataaccaacc cgaaatgact caagaaaaat ttaaacctag 60

ctttcttgat gagacaaaaa ctctctttag aaccggcgat ttaggcaagc aaactgctcc 120

gggtatcatt gagtttatgg gacgaaaaga taatcaagtt aaggtcaatg gttatcgaat 180

tgaccccgga gaaattgaat atcaattgac tcgttatgct cccattgaaa gagcgattgt 240

tttacccgtt caagttaata atcaaactca attatctgct tactgtcaaa cagacaaaac 300

-continued

tctagaaatt gctgagattc gagaattact tgccaaattt ttaccagttt atatgattcc 360

gagttacttt atttttttaa agcaattccc cttaactcga catggaaaac ttgacctgca 420

ctccctgaga gaactcagag aaactggtaa atctctggtg aattctaatt acgttgcacc 480

ccggaattat ttagaatcca atctcgttag tatctgggaa aaaattctct ctaaacatcc 540

tatcggtatt tttgataact tctttgaaat tggcggtcat tctctactct tatcaagggt 600

tgtaacccgg gttcataaag aactaaatgt atccgtaaaa ttagctgact tctttaaagt 660

tccaaccgtt gctggattgg cgactttaat ctcccagact caatacaatt atcaagaacc 720

catttcggca attcccccccc aaaaatctta tccgatgtct catggtcagc gt 772

<210> SEQ ID NO 17
<211> LENGTH: 772
<212> TYPE: DNA
<213> ORGANISM: Microcystis Miz25

<400> SEQUENCE: 17

cgccctagca tcgggttatc ataaccaacc cgaaatgact caagaaaaat ttaaacctag 60

ctttcttgat gagacaaaaa ctctctttag aaccggcgat ttaggcaagc aaactgctcc 120

cggtatcatt gagtttatgg gacgaaaaga taatcaagtt aaggtcaatg gttatcgaat 180

tgaccccgga gaaattgaat atcaattgac tcgctatgct cccattgaaa gagcgattgt 240

tttacccgtt caagttaata atcaaactca attatctgct tactgtcaaa cagacaaaac 300

tctagaaatt gctgagattc gagaattact tgccaaattt ttaccagttt atatgattcc 360

gagttacttt atttttttaa agcaattccc cttaactcga catgggaaac ttgacctgca 420

ctccctgaga gaactcagag aaactggtaa atctctggtg aattctaatt atgttgcacc 480

ccggaattat ttagaatcca atctcgttag tatctgggaa aaaattctct ctaaacatcc 540

tatcggtatt tttgataact tctttgaaat tggcggtcat tctctactct tatcaagggt 600

tgtaacccgg gttcataaag aactaaatgt atccgtaaaa ttggctgact tctttaaagt 660

tccaaccatt gctggattgg cgactttaat ctcccagact caatacaatt atcaagaacc 720

catttcggca attcccccccc aaaaatcctа tccgatgtct catggtcagc gt 772

<210> SEQ ID NO 18
<211> LENGTH: 772
<212> TYPE: DNA
<213> ORGANISM: Microcystis MizM5

<400> SEQUENCE: 18

cgccctagca tcgggttatc ataaccaacc cgaaatgact caagaaaaat ttaaacctag 60

ctttcttgat gagacaaaaa ctctctttag aaccggcgat ttaggcaagc aaactgctcc 120

gggtatcatt gagtttatgg gacgaaaaga taatcaagtt aaggtcaatg gttatcgaat 180

tgatcccgga gaaattgaat atcaattgac tcgttatgct cccattgaaa gagcgattgt 240

tttacccatt caagtgaata atcaaactca attatctgct tactgtcaaa cagaaaaaac 300

tctagaaatt gctgagattc gagaatttct tgccaagttt ttaccagttt atatgattcc 360

cagttacttt atttttttaa agcaattccc tttaactcga catgggaaac ttgacctgca 420

ctccctgaga caactcagag aaaccggtaa atatctggtt aattctaatt atgttgcacc 480

ccgtaatcat ttagaatcca atctcgttag tatctgggaa aaaattctct ctaaacatcc 540

-continued

tatcggtatt tttgataact tctttgaaat tggcggtcat tctctactct tatcaagggt 600

tgtaacccgg gttcataaag aactaaatgt atccgtaaaa ttggctgact tctttaaagt 660

tccaaccatt gctggattgg cgactttaat ctcccagact caatacaatt atcaagaacc 720

catttcggca attccccccc aaaaatctta tccgatgtct catggtcagc gt 772

<210> SEQ ID NO 19
<211> LENGTH: 772
<212> TYPE: DNA
<213> ORGANISM: Microcystis Mnies102

<400> SEQUENCE: 19

cgccctagca tcgggttatc ataaccaacc cgaaatgact caagaaaaat ttaaacctag 60

ctttcttgat gagacaaaaa ctctctttag aaccggcgat ttaggcaagc aaactgctcc 120

cggtatcatt gagtttatgg gacgaaaaga taatcaagtt aaggtcaatg gttatcgaat 180

tgaccccgga gaaattgaat atcaattgac tcgttatgct cccattgaaa gagcgattgt 240

tttacccgtt caagtgaata atcaaactca attatctgct tactgtcaaa cagacaaaac 300

tctagaaatt gctgagattc gagaatttct tgccaagttt ttgccagttt atatgattcc 360

cagttacttt atttttttaa agcaattccc cttaactcga catgggaaac ttgacctgca 420

ctccctgaga gaactcaaag aaaccagtaa atctctggtt aattctaatt atgttgcacc 480

ccgtaatcat ttagaatcca atctcgttag tatctgggaa aaaattctct ctaaacatcc 540

tatcggtatt tttgataact tttttgaaat tggtggtcat tctctactct tatcaagggt 600

tgtaacccgg gttcataaag aactaaatgt atccgtaaaa ttagctgact tctttaaagt 660

tccaaccgtt gctggattgg cgactttaat ctcccagact caatacaatt atcaagaacc 720

catttcggca attccccccc aaaaatctta tccgatgtct catggtcagc gt 772

<210> SEQ ID NO 20
<211> LENGTH: 772
<212> TYPE: DNA
<213> ORGANISM: Microcystis MniesA89

<400> SEQUENCE: 20

cgccctagca tcgggttatc ataaccaacc cgaaatgact caagaaaaat ttaaacctag 60

ctttcttgat gagacaaaaa ctctctttag aaccggcgat ttaggcaagc aaactgctcc 120

cggtatcatt gagtttatgg gacgaaaaga taatcaagtt aaggtcaatg gttatcgaat 180

tgaccccgga gaaattgaat atcaattgac tcgttatgct cccattgaaa gagcgattgt 240

tttacccgtt caagtgaata atcaaactca attatctgct tactgtcaaa cagacaaaac 300

tctagaaatt gctgagattc gagaatttct tgccaagttt ttgccagttt atatgattcc 360

cagttacttt atttttttaa agcaattccc cttaactcga catgggaaac ttgacctgca 420

ctccctgaga gaactcaaag aaaccagtaa atctctggtt aattctaatt atgttgcacc 480

ccgtaatcat ttagaatcca atctcgttag tatctgggaa aaaattctct ctaaacatcc 540

tatcggtatt tttgataact tctttgaaat tggcggtcat tctctactct tatcaagggt 600

tgtaacccgg gttcataaag aactaaatgt atccgtaaaa ttggctgact tctttaaagt 660

tccaaccatt gctggattgg cgactttaat ctcccagact caatacaatt atcaagaacc 720

catttcggca attccccccc aaaaatccta tccgatgtct catggtcagc gt 772

-continued

<210> SEQ ID NO 21
<211> LENGTH: 772
<212> TYPE: DNA
<213> ORGANISM: Microcystis consensus

<400> SEQUENCE: 21

cgccctagca tcgggttatc ataaccaacc cgaaatgact caagaaaaat ttaaacctag 60

ctttcttgat gagacaaaaa ctctctttag aaccggcgat ttaggcaagc aaactgctcc 120

cggtatcatt gagtttatgg gacgaaaaga taatcaagtt aaggtcaatg gttatcgaat 180

tgaccccgga gaaattgaat atcaattgac tcgttatgct cccattgaaa gagcgattgt 240

tttacccgtt caagttaata atcaaactca attatctgct tactgtcaaa cagacaaaac 300

tctagaaatt gctgagattc gagaattact tgccaaattt ttaccagttt atatgattcc 360

gagttacttt attttttaa agcaattccc cttaactcga catggaaaac ttgacctgca 420

ctccctgaga gaactcagag aaactggtaa atctctggtg aattctaatt acgttgcacc 480

ccggaattat ttagaatcca atctcgttag tatctgggaa aaaattctct ctaaacatcc 540

tatcggtatt tttgataact tctttgaaat tggcggtcat tctctactct tatcaagggt 600

tgtaacccgg gttcataaag aactaaatgt atccgtaaaa ttagctgact tctttaaagt 660

tccaaccgtt gctggattgg cgactttaat ctcccagact caatacaatt atcaagaacc 720

catttcggca attcccccc aaaaatctta tccgatgtct catggtcagc gt 772

<210> SEQ ID NO 22
<211> LENGTH: 769
<212> TYPE: DNA
<213> ORGANISM: Planktothrix P49

<400> SEQUENCE: 22

agccttagca tcgggttatc acaatcttcc tcaaatcaca aaagaaaaat ttaaacctgg 60

ctttttaat cagaaaacaa cgatgtttag aaccggggat ttagggaaac aaactgctcc 120

cggtgtgatt gaatttatgg gcagaaaaga caatcaagtt aaggtaaatg gctatcgtat 180

cgaccccgaa gaaattgaat atcaacttaa tcgttatcct cagattgaga gagctattat 240

tctaccgata tcagtcaata atcaaactca attatcagcc tattgtcaaa ccgataaaca 300

gatagaaatt tctgaaatca gagaatttct agctaatttt ctgccagttt acatgattcc 360

tagttacttt attttcttaa agcaattccc cctaactaaa cacggcaaac ttgacttaaa 420

ctcactgatt gcactcaatg aaaccgggaa atctacccag gtaaattatg ttgcaccgcg 480

taataattta gagtcaaacc tagttagaat ctgggaaaag attctgacca aacatcccat 540

cggtattttt gataacttct ttgaaattgg cggacattct ctgatgcttt cgagaatcgt 600

aacccacgtt cataaagaat taaatgtatc ggtaaaattg gctgacttct ttaaagttcc 660

taccattgcc ggattagccg ttttagtctc taaaactgaa tatgattatc aagaacccat 720

tcccacaatt cctctgcaaa aatcctatcc gatgtcccat gggcaacgt 769

<210> SEQ ID NO 23
<211> LENGTH: 769
<212> TYPE: DNA
<213> ORGANISM: Planktothrix P97

<400> SEQUENCE: 23

-continued

agccttagca tcgggttatc acaatcttcc tcaaatcaca aaagaaaaat ttaaacctgg 60

ctttttgat cagaaaacaa cgatgtttag aaccggggat ttagggaaac aaactgctcc 120

cggtgtgatt gaatttatgg gcagaaaaga caatcaagtt aaggtaaatg gctatcgtat 180

cgaccccgaa gaaattgaat atcaacttaa tcgttatcct cagattgaga gagctattat 240

tctaccgata tcagtcaata atcaaactca attatcagcc tattgtcaaa cagataaaca 300

gatagaaatt tctgaaatta gagaatttct agctaatttt ctgccagttt acatgattcc 360

tagttacttt attttcttaa agcaatttcc cctaactaaa cacggcaaac ttgacttaaa 420

ctcaatgatt gcactcaatg aaaccgggaa atctacccaa gtaaattatg ttgcaccgcg 480

taataattta gagtcaaacc tagttagaat ctgggaaaag attctgacca aacatcccat 540

cggtattttt gataacttct ttgaaattgg cggacattct ctgatgcttt cgagaatcgt 600

aacccacgtt cataaagaat taaatgtatc ggtaaaattg gctgacttct ttaaagttcc 660

taccattgcc ggattagccg ttttagtctc taaaactgaa tatgattatc aagaacccat 720

tcccacaatt cctctgcaaa aatcctatcc gatgtcccat gggcaacgt 769

<210> SEQ ID NO 24
<211> LENGTH: 769
<212> TYPE: DNA
<213> ORGANISM: Planktothrix P126-8

<400> SEQUENCE: 24

agccttagca tcgggttatc acaatcttcc tcaaatcaca aaagaaaaat ttaaacctgg 60

ctttttgat cagaaaacaa cgatgtttag aaccggggat ttagggaaac aaactgctcc 120

cggtgtgatt gaatttatgg gcagaaaaga caatcaagtt aaggtaaatg gctatcgtat 180

cgaccccgaa gaaattgaat atcaacttaa tcgttatcct cagattgaga gagctattat 240

tctaccgata tcagtcaata atcaaactca attatcagcc tattgtcaaa cagataaaca 300

gatagaaatt tctgaaatta gagaatttct agctaatttt ttgccagttt acatgattcc 360

tagttacttt attttcttaa agcaatttcc cctaactaaa cacggcaaac ttgacttaaa 420

ctcaatgatt gcactcaatg aaaccgggaa atctacccaa gtaaattatg ttgcaccgcg 480

taataattta gagtcaaacc tagttagaat ctgggaaaag attctaacca aacatcccat 540

cggtattttt gataacttct ttgaaattgg cggacattct ctgatgcttt cgagaatcgt 600

aacccacgtt cataaagaat taaatgtatc ggtaaaattg gctgacttct ttaaagttcc 660

taccattgcc ggattagccg ttttagtctc taaaactgaa tatgattatc aagaacccat 720

ccccacaatt cctctgcaaa aatcctatcc gatgtcccat gggcaacgt 769

<210> SEQ ID NO 25
<211> LENGTH: 769
<212> TYPE: DNA
<213> ORGANISM: Planktothrix P127

<400> SEQUENCE: 25

agccttagca tcgggttatc acaatcttcc tcaaatcaca aaagaaaaat ttaaacctgg 60

ctttttgat cagaaaacaa cgatgtttag aaccggggat ttagggaaac aaactgctcc 120

cggtgtgatt gaatttatgg gcagaaaaga caatcaagtt aaggtaaatg gctatcgtat 180

cgaccccgaa gaaattgaat atcaacttaa tcgttatcct cagattgaga gagctattat 240

-continued

tctaccgata tcagtcaata atcaaactca attatcagcc tattgtcaaa cagataaaca 300

gatagaaatt tctgaaatta gagaatttct agctaatttt ttgccagttt acatgattcc 360

tagttacttt attttcttaa agcaatttcc cctaactaaa cacggcaaac ttgacttaaa 420

ctcaatgatt gcactcaatg aaaccgggaa atctacccaa gtaaattatg ttgcaccgcg 480

taataattta gagtcaaacc tagttagaat ctgggaaaag attctaacca aacatcccat 540

cggtattttt gataacttct ttgaaattgg cggacattct ctgatgcttt cgagaatcgt 600

aacccacgtt cataaagaat taaatgtatc ggtaaaattg gctgacttct ttaaagttcc 660

taccattgcc ggattagccg ttttagtctc taaaactgaa tatgattatc aagaacccat 720

ccccacaatt cctctgcaaa aatcctatcc gatgtcccat gggcaacgt 769

<210> SEQ ID NO 26
<211> LENGTH: 769
<212> TYPE: DNA
<213> ORGANISM: Planktothrix P128-R

<400> SEQUENCE: 26

agccttagca tcgggttatc acaatcttcc tcaaatcaca aaagaaaaat ttaaacctgg 60

ctttttgat cagaaaacaa cgatgtttag aaccggggat ttagggaaac aaactgctcc 120

cggtgtgatt gaatttatgg gcagaaaaga caatcaagtt aaggtaaatg gctatcgtat 180

cgaccccgaa gaaattgaat atcaacttaa tcgttatcct cagattgaga gagctattat 240

tctaccgata tcagtcaata atcaaactca attatcagcc tattgtcaaa cagataaaca 300

gatagaaatt tccgaaatta gagaatttct agctaatttt ctgccagttt acatgattcc 360

tagttacttt attttcttaa agcaatttcc cctaactaaa cacggcaaac ttgacttaaa 420

ctcaatgatt gcactcaatg aaaccgggaa atctacccaa gtaaattatg ttgcaccgcg 480

taataattta gagtcaaacc tagttagaat ctgggaaaag attctaacca aacatcccat 540

cggtattttt gataacttct ttgaaattgg cggacattct ctgatgcttt cgagaatcgt 600

aacccacgtt cataaagaat taaatgtatc ggtaaaattg gctgacttct ttaaagttcc 660

taccattgcc ggattagccg ttttagtctc taaaactgaa tatgattatc aagaacccat 720

ccccacaatt cctctgcaaa aatcctatcc gatgtcccat gggcaacgt 769

<210> SEQ ID NO 27
<211> LENGTH: 769
<212> TYPE: DNA
<213> ORGANISM: Oscillatoria O213

<400> SEQUENCE: 27

agccttagca tcgggttatc acaatcttcc tcaaatcaca aaagaaaaat ttaaacctgg 60

ctttttaat cagaaaacaa cgatgtttag aaccggggat ttagggaaac aaactgctcc 120

cggtgtgatt gaatttatgg gcagaaaaga caatcaagtt aaggtaaatg gctatcgtat 180

cgaccccgaa gaaattgaat atcaacttaa tcgttatcct cagattgaga gagctattat 240

tctaccgata tcagtcaata atcaaactca attatcagcc tattgtcaaa ccgataaaca 300

gatagaaatt tctgaaatca gagaatttct agctaatttt ctgccagttt acatgattcc 360

tagttacttt attttcttaa agcaattccc cctaactaaa cacggcaaac ttgacttaaa 420

ctcactgatt gcactcaatg aaaccgggaa atctacccag gtaaattatg ttgcaccgcg 480

-continued

taataattta gagtcaaacc tagttagaat ctgggaaaag attctgacca aacatcccat 540

cggtattttt gataacttct ttgaaattgg cggacattct ctgatgcttt cgagaatcgt 600

aacccacgtt cataaagaat taaatgtatc ggtaaaattg gctgacttct ttaaagttcc 660

taccattgcc ggattagccg ttttagtctc taaaactgaa tatgattatc aagaacccat 720

tcccacaatt cctctgcaaa aatcctatcc gatgtcccat gggcaacgt 769

<210> SEQ ID NO 28
<211> LENGTH: 769
<212> TYPE: DNA
<213> ORGANISM: Oscillatoria O226

<400> SEQUENCE: 28

agccttagca tcgggttatc acaatcttcc tcaaatcaca aaagaaaaat ttaaacctgg 60

ctttttttaat cagaaaacaa cgatgtttag aaccggggat ttagggaaac aaactgctcc 120

cggtgtgatt gaatttatgg gcagaaaaga caatcaagtt aaggtaaatg gctatcgtat 180

cgaccccgaa gaaattgaat atcaacttaa tcgttatcct cagattgaga gagctattat 240

tctaccgata tcagtcaata atcaaactca attatcagcc tattgtcaaa ccgataaaca 300

gatagaaatt tctgaaatca gagaatttct agctaatttt ctgccagttt acatgattcc 360

tagttacttt attttcttaa agcaattccc cctaactaaa cacggcaaac ttgacttaaa 420

ctcactgatt gcactcaatg aaaccgggaa atctacccag gtaaattatg ttgcaccgcg 480

taataattta gagtcaaacc tagttagaat ctgggaaaag attctgacca aacatcccat 540

cggtattttt gataacttct ttgaaattgg cggacattct ctgatgcttt cgagaatcgt 600

aacccacgtt cataaagaat taaatgtatc ggtaaaattg gctgacttct ttaaagttcc 660

taccattgcc ggattagccg ttttagtctc taaaactgaa tatgattatc aagaacccat 720

tcccacaatt cctctgcaaa aatcctatcc gatgtcccat gggcaacgt 769

<210> SEQ ID NO 29
<211> LENGTH: 769
<212> TYPE: DNA
<213> ORGANISM: Oscillatoria/Planktothrix consensus

<400> SEQUENCE: 29

agccttagca tcgggttatc acaatcttcc tcaaatcaca aaagaaaaat ttaaacctgg 60

ctttttttgat cagaaaacaa cgatgtttag aaccggggat ttagggaaac aaactgctcc 120

cggtgtgatt gaatttatgg gcagaaaaga caatcaagtt aaggtaaatg gctatcgtat 180

cgaccccgaa gaaattgaat atcaacttaa tcgttatcct cagattgaga gagctattat 240

tctaccgata tcagtcaata atcaaactca attatcagcc tattgtcaaa cagataaaca 300

gatagaaatt tctgaaatta gagaatttct agctaatttt ctgccagttt acatgattcc 360

tagttacttt attttcttaa agcaatttcc cctaactaaa cacggcaaac ttgacttaaa 420

ctcaatgatt gcactcaatg aaaccgggaa atctacccaa gtaaattatg ttgcaccgcg 480

taataattta gagtcaaacc tagttagaat ctgggaaaag attctgacca aacatcccat 540

cggtattttt gataacttct ttgaaattgg cggacattct ctgatgcttt cgagaatcgt 600

aacccacgtt cataaagaat taaatgtatc ggtaaaattg gctgacttct ttaaagttcc 660

taccattgcc ggattagccg ttttagtctc taaaactgaa tatgattatc aagaacccat 720

-continued

tcccacaatt cctctgcaaa aatcctatcc gatgtcccat gggcaacgt 769

<210> SEQ ID NO 30
<211> LENGTH: 769
<212> TYPE: DNA
<213> ORGANISM: Nodularia Nod-HEMclone

<400> SEQUENCE: 30

agcattagca tcaggttatc acaacctccc ccaaatcacc gcagaaaaat ttcaacctag 60

ctttatgact gagggaaaaa ctatctttag aaccggagat ttaggtaaac aaattgcccc 120

aggcgtgatt gaatttcttg gtcgtaaaga taatcaagtt aaggtgaatg gctatcgtat 180

agatccagga gaaattgaat accaactcag ccgccattct caaattgaga gagcaatcgt 240

attgcctact aatgtagata atcaaaccca gttatcagcc tattgtaaaa ctgagtcaga 300

catagaaatt tccgaaattc gagaatttct atcgaacttt ttgcccgttt acatgattcc 360

tactttcttt atcttcttaa agcaatttcc cttaaccaga catgggaaaa ttgatttgcg 420

atccctggct gaattcaagg gaataggtaa cttaacacag ttagcgtata ctgcaccgcg 480

caataattta gagtccaagc tcgtacatat ttgggaaaaa attctcacca aacaacccat 540

tggcattttt gataacttct ttgaaattgg tggacactca ctgctgcttt ccagagtggt 600

aactcacgtt cataaagaat taaatgtgtt ggtaaaattg gctgaattct ttaaagttcc 660

cacaatcgcc ggattagcag ctttagtatc taaaacccaa tatgactatc aagaacccat 720

accagcaata actcagcaaa cgtcttatcc tatgtctcat gggcaacgc 769

<210> SEQ ID NO 31
<211> LENGTH: 769
<212> TYPE: DNA
<213> ORGANISM: Nodularia Nod-BY1clone

<400> SEQUENCE: 31

agcattagca gcaggttatc acaacctccc ccaaatcacc gcagaaaaat ttcaacctag 60

ctttatgact gagggaaaaa ctatctttag aaccggagat ttaggtaaac aaattgcccc 120

aggcgtgatt gaatttcttg gtcgtaaaga taatcaagtt aaggtgaatg gctatcgcat 180

agatccagga gaaattgaat accaactcag ccgccattct caaattgaga gagcaatcgt 240

attgcctact aatgtagata atcaaaccca gttatcagcc tattgtaaaa ctgagtcaga 300

catagaaatt tccgaaattc gagaatttct atcgaacttt ttgcccgttt acatgattcc 360

tactttcttt atcttcttaa agcaatttcc cttaaccaga catgggaaaa ttgatttgcg 420

atccctggct gaattcaagg gaataggtaa cttaacacag ttagcgtata ctgcaccgcg 480

caataattta gagtccaagc tcgtacatat ttgggaaaaa attctcacca aacaacccat 540

tggcattttt gataacttct ttgaaattgg tggacactca ctgctgcttt ccagagtggt 600

aactcacgtt cataaagaat taaatgtgtt ggtaaaattg gctgaattct ttaaagttcc 660

cacaatcgcc ggattagcag ctttagtatc taaaacccaa tatgactatc aagaacccat 720

accagcaata actcagcaaa cgtcttatcc tatgtctcat gggcaacgc 769

<210> SEQ ID NO 32
<211> LENGTH: 769
<212> TYPE: DNA
<213> ORGANISM: Nodularia Nod-F81clone

-continued

<400> SEQUENCE: 32

agcattagca tcaggttatc acaacctccc ccaaatcacc gcagaaaaat ttcaacctag 60

ctctatgact gagggaaaaa ctatctttag aaccggagat ttaggtaaac aaattgcccc 120

aggcgtgatt gaatttcttg gtcgtaaaga taatcaagtt aaggtgaatg gctatcgtat 180

agatccagga gaaattgaat accaactcag ccgccattct caaattgaga gagcaatcgt 240

attgcctact aatgtagata atcaaaccca gttatcagcc tattgtaaaa ctgagtcaga 300

catagaaatt tccgaaattc gagaatttct atcgaacttt ttgcccgttt atatgattcc 360

tactttcttt atcttcttaa agcaatttcc cttaaccaga catgggaaaa ttgatttgcg 420

atccctggct gaattcaagg gaataggtaa cttaacacag ttagcgtata ctgcaccgcg 480

caataattta gagtccaagc tcgtacatat ttgggaaaaa attctcacca aacaacccat 540

tggcattttt gataacttct ttgaaattgg tggacactca ctgctgcttt ccagagtggt 600

aactcacgtt cataaagaat taaatgtgtt ggtaaaattg gctgaattct ttaaagttcc 660

cacaatcgcc ggattagcag ctttagtatc taaaacccaa tatgactatc aagaacccat 720

accagcaata actcagcaaa cgtcttatcc tatgtctcat gggcaacgc 769

<210> SEQ ID NO 33
<211> LENGTH: 769
<212> TYPE: DNA
<213> ORGANISM: Nodularia consensus

<400> SEQUENCE: 33

agcattagca tcaggttatc acaacctccc ccaaatcacc gcagaaaaat ttcaacctag 60

ctttatgact gagggaaaaa ctatctttag aaccggagat ttaggtaaac aaattgcccc 120

aggcgtgatt gaatttcttg gtcgtaaaga taatcaagtt aaggtgaatg gctatcgtat 180

agatccagga gaaattgaat accaactcag ccgccattct caaattgaga gagcaatcgt 240

attgcctact aatgtagata atcaaaccca gttatcagcc tattgtaaaa ctgagtcaga 300

catagaaatt tccgaaattc gagaatttct atcgaacttt ttgcccgttt acatgattcc 360

tactttcttt atcttcttaa agcaatttcc cttaaccaga catgggaaaa ttgatttgcg 420

atccctggct gaattcaagg gaataggtaa cttaacacag ttagcgtata ctgcaccgcg 480

caataattta gagtccaagc tcgtacatat ttgggaaaaa attctcacca aacaacccat 540

tggcattttt gataacttct ttgaaattgg tggacactca ctgctgcttt ccagagtggt 600

aactcacgtt cataaagaat taaatgtgtt ggtaaaattg gctgaattct ttaaagttcc 660

cacaatcgcc ggattagcag ctttagtatc taaaacccaa tatgactatc aagaacccat 720

accagcaata actcagcaaa cgtcttatcc tatgtctcat gggcaacgc 769

<210> SEQ ID NO 34
<211> LENGTH: 769
<212> TYPE: DNA
<213> ORGANISM: Nostoc 152

<400> SEQUENCE: 34

agcattagcc gcaggttatc ataatcttcc cgacatcact acagaaaaat ttcaacccag 60

cttgataagt gagggaaaaa ctctctttag aacgggagat ttaggtaaac aaactgctcc 120

aggtgtcatt gaatttattg ggcgtaaaga taatcaagta aaggtgaatg gttatcgtat 180

-continued

agatccagga gaaattgaat atcaactcag ccgtcatgct cagattgaaa gagcgattat 240

attgcctatc aatgtagata atcaaactca attatctgct tattgtcaaa ctgataaaga 300

catcgaaatt gctgaaatta gagaatttct ctctaaattt ttgccagttt atatgattcc 360

tacttccttt atcttcctca agcaatttcc tctaaccaga catggcaaaa ttgacttgcg 420

atcgcttgct gaactccaag gaatcggtaa gttaacacag gcagattata ctgcaccgcg 480

caatgattta gagtccaagc tagtaaagat ttgggaaaaa attctcacca cacatcccat 540

cggcattttt gataacttct ttgaaattgg tggacactcg ctgctgcttt cgagagtggt 600

aacttacgtt cataaagaat taaatgtgtt agtcaaattg gctgactttt ttaaagttcc 660

caccatagcc ggattagcag ctttagtagc taaaacccaa tacgattatc aagaacccat 720

acctgcaata attcagcaaa aatcttatcc catgtctcat gggcaacgc 769

<210> SEQ ID NO 35
<211> LENGTH: 769
<212> TYPE: DNA
<213> ORGANISM: Anabaena consensus McyE

<400> SEQUENCE: 35

agcattagca gttggttatc gtaatcttcc tgaaattacc agagaaaaat ttcaacccaa 60

ctttttaaat tcagaaaaaa ttctctttag aacgggagat ttaggtaaac aaattgctcc 120

aggtgtgatt gaatttatag ggcgaaaaga taatcaagtt aaggtcaatg gctatcgtgt 180

agatccagga gaamttgaat accaaattag ccgctatgcc gagattgaga aagcaattgt 240

cttacctata gaggtaaata accaaattca attatctgct tattgtcaaa ctgataaaga 300

tataaaawtt tctgaaatca gagaattttt agctaaatat tygccagttt acatgattcc 360

tagttccttt atcttcttaa agcaatttcc cttaactaaa catggcaaay ttgacttgcg 420

atcgctcrtc gctctcaagc cracagatca aytaacacaa gtctcttata ctgcaccgcg 480

taatacttta gaatcaaagc tagtccatat ttgggaaaaa attctcacta aacatcccat 540

tggaattttt gataactttt ttgaaatcgg cggacactct ctgctccttt ctagagtagt 600

cactcacgtc cataaagaat taaatgtatt agttaaatta gctgatttct tcaaagttcc 660

cacaattctt ggattagcag ctttaatatc taaagctcaa tctaactatc aagaacccat 720

accagcaata actcaacaag aatcttatcc catgtctcat ggacaacgc 769

<210> SEQ ID NO 36
<211> LENGTH: 772
<212> TYPE: DNA
<213> ORGANISM: Microcystis consensus McyE

<400> SEQUENCE: 36

cgccctagca tcgggttatc ataaccaacc cgaaatgact caagaaaaat ttaaacctag 60

ctttcttgat gagacaaaaa ctctctttag aaccggcgat ttaggcaagc aaactgctcc 120

sggtatcatt gagtttatgg gacgaaaaga taatcaagtt aaggtcaatg gttatcgaat 180

tgaccccgga gaaattgaat atcaattgac tcgttatgct cccattgaaa gagcgattgt 240

tttacccgtt caagtkaata atcaaactca attatctgct tactgtcaaa cagacaaaac 300

tctagaaatt gctgagattc gagaattwct tgccaarttt ttaccagttt atatgattcc 360

sagttacttt attttttaa agcaattccc cttaactcga catggraaac ttgacctgca 420

-continued

ctccctgaga gaactcagag aaacyggtaa atctctggtk aattctaatt aygttgcacc 480

ccgkaatyat ttagaatcca atctcgttag tatctgggaa aaaattctct ctaaacatcc 540

tatcggtatt tttgataact tctttgaaat tggcggtcat tctctactct tatcaagggt 600

tgtaacccgg gttcataaag aactaaatgt atccgtaaaa ttrgctgact tctttaaagt 660

tccaaccrtt gctggattgg cgactttaat ctcccagact caatacaatt atcaagaacc 720

catttcggca attccccccc aaaaatcyta tccgatgtct catggtcagc gt 772

<210> SEQ ID NO 37
<211> LENGTH: 769
<212> TYPE: DNA
<213> ORGANISM: Nodularia consensus McyE

<400> SEQUENCE: 37

agcattagca kcaggttatc acaacctccc ccaaatcacc gcagaaaaat ttcaacctag 60

ctytatgact gagggaaaaa ctatctttag aaccggagat ttaggtaaac aaattgcccc 120

aggcgtgatt gaatttcttg gtcgtaaaga taatcaagtt aaggtgaatg gctatcgyat 180

agatccagga gaaattgaat accaactcag ccgccattct caaattgaga gagcaatcgt 240

attgcctact aatgtagata atcaaaccca gttatcagcc tattgtaaaa ctgagtcaga 300

catagaaatt tccgaaattc gagaatttct atcgaacttt ttgcccgttt ayatgattcc 360

tactttcttt atcttcttaa agcaatttcc cttaaccaga catgggaaaa ttgatttgcg 420

atccctggct gaattcaagg gaataggtaa cttaacacag ttagcgtata ctgcaccgcg 480

caataattta gagtccaagc tcgtacatat ttgggaaaaa attctcacca aacaacccat 540

tggcattttt gataacttct ttgaaattgg tggacactca ctgctgcttt ccagagtggt 600

aactcacgtt cataaagaat taaatgtgtt ggtaaaattg gctgaattct ttaaagttcc 660

cacaatcgcc ggattagcag ctttagtatc taaaacccaa tatgactatc aagaacccat 720

accagcaata actcagcaaa cgtcttatcc tatgtctcat gggcaacgc 769

<210> SEQ ID NO 38
<211> LENGTH: 769
<212> TYPE: DNA
<213> ORGANISM: Nostoc consensus McyE

<400> SEQUENCE: 38

agcattagcc gcaggttatc ataatcttcc cgacatcact acagaaaaat ttcaacccag 60

cttgataagt gagggaaaaa ctctctttag aacgggagat ttaggtaaac aaactgctcc 120

aggtgtcatt gaatttattg ggcgtaaaga taatcaagta aaggtgaatg gttatcgtat 180

agatccagga gaaattgaat atcaactcag ccgtcatgct cagattgaaa gagcgattat 240

attgcctatc aatgtagata atcaaactca attatctgct tattgtcaaa ctgataaaga 300

catcgaaatt gctgaaatta gagaatttct ctctaaattt ttgccagttt atatgattcc 360

tacttccttt atcttcctca agcaatttcc tctaaccaga catggcaaaa ttgacttgcg 420

atcgcttgct gaactccaag gaatcggtaa gttaacacag gcagattata ctgcaccgcg 480

caatgattta gagtccaagc tagtaaagat ttgggaaaaa attctcacca cacatcccat 540

cggcattttt gataacttct ttgaaattgg tggacactcg ctgctgcttt cgagagtggt 600

aacttacgtt cataaagaat taaatgtgtt agtcaaattg gctgactttt ttaaagttcc 660

-continued

caccatagcc ggattagcag ctttagtagc taaaacccaa tacgattatc aagaacccat 720

acctgcaata attcagcaaa aatcttatcc catgtctcat gggcaacgc 769

<210> SEQ ID NO 39
<211> LENGTH: 769
<212> TYPE: DNA
<213> ORGANISM: Oscillatoria/Planktothrix consensus McyE

<400> SEQUENCE: 39

agccttagca tcgggttatc acaatcttcc tcaaatcaca aaagaaaaat ttaaacctgg 60

ctttttttrat cagaaaacaa cgatgtttag aaccggggat ttagggaaac aaactgctcc 120

cggtgtgatt gaatttatgg gcagaaaaga caatcaagtt aaggtaaatg gctatcgtat 180

cgaccccgaa gaaattgaat atcaacttaa tcgttatcct cagattgaga gagctattat 240

tctaccgata tcagtcaata atcaaactca attatcagcc tattgtcaaa cmgataaaca 300

gatagaaatt tctgaaatya gagaatttct agctaatttt ctgccagttt acatgattcc 360

tagttacttt attttcttaa agcaattycc cctaactaaa cacggcaaac ttgacttaaa 420

ctcamtgatt gcactcaatg aaaccgggaa atctacccar gtaaattatg ttgcaccgcg 480

taataattta gagtcaaacc tagttagaat ctgggaaaag attctracca aacatcccat 540

cggtattttt gataacttct ttgaaattgg cggacattct ctgatgcttt cgagaatcgt 600

aacccacgtt cataaagaat taaatgtatc ggtaaaattg gctgacttct ttaaagttcc 660

taccattgcc ggattagccg ttttagtctc taaaactgaa tatgattatc aagaacccat 720

ycccacaatt cctctgcaaa aatcctatcc gatgtcccat gggcaacgt 769

<210> SEQ ID NO 40
<211> LENGTH: 21
<212> TYPE: DNA
<213> ORGANISM: artificial sequence
<220> FEATURE:
<223> OTHER INFORMATION: discriminating probe; chemically synthetized

<400> SEQUENCE: 40

accaaattag ccgctatgcc g 21

<210> SEQ ID NO 41
<211> LENGTH: 28
<212> TYPE: DNA
<213> ORGANISM: artificial sequence
<220> FEATURE:
<223> OTHER INFORMATION: discriminating probe; chemically synthetized

<400> SEQUENCE: 41

tctactctta tcaagggttg taacccgg 28

<210> SEQ ID NO 42
<211> LENGTH: 21
<212> TYPE: DNA
<213> ORGANISM: artificial sequence
<220> FEATURE:
<223> OTHER INFORMATION: discriminating probe; chemically synthetized

<400> SEQUENCE: 42

atttgcgatc cctggctgaa t 21

<210> SEQ ID NO 43

-continued

<211> LENGTH: 26
<212> TYPE: DNA
<213> ORGANISM: artificial sequence
<220> FEATURE:
<223> OTHER INFORMATION: discriminating probe; chemically synthetized

<400> SEQUENCE: 43

acttctttga aattggtgga cactcg 26

<210> SEQ ID NO 44
<211> LENGTH: 25
<212> TYPE: DNA
<213> ORGANISM: artificial sequence
<220> FEATURE:
<223> OTHER INFORMATION: discriminating probe; chemically synthetized

<400> SEQUENCE: 44

gatgtttaga accggggatt taggg 25

<210> SEQ ID NO 45
<211> LENGTH: 32
<212> TYPE: DNA
<213> ORGANISM: artificial sequence
<220> FEATURE:
<223> OTHER INFORMATION: discriminating probe;chemically synthetized

<400> SEQUENCE: 45

aacttaatcg ttatcctcag attgagagag ct 32

<210> SEQ ID NO 46
<211> LENGTH: 33
<212> TYPE: DNA
<213> ORGANISM: artificial sequence
<220> FEATURE:
<223> OTHER INFORMATION: common probe;chemically synthetized

<400> SEQUENCE: 46

agattgagaa agcaattgtc ttacctatag agg 33

<210> SEQ ID NO 47
<211> LENGTH: 37
<212> TYPE: DNA
<213> ORGANISM: artificial sequence
<220> FEATURE:
<223> OTHER INFORMATION: common probe; chemically synthetized

<400> SEQUENCE: 47

gttcataaag aactaaatgt atccgtaaaa ttrgctg 37

<210> SEQ ID NO 48
<211> LENGTH: 32
<212> TYPE: DNA
<213> ORGANISM: artificial sequence
<220> FEATURE:
<223> OTHER INFORMATION: common probe; chemically synthetized

<400> SEQUENCE: 48

tcaagggaat aggtaactta acacagttag cg 32

<210> SEQ ID NO 49
<211> LENGTH: 28
<212> TYPE: DNA
<213> ORGANISM: artificial sequence
<220> FEATURE:

-continued

<223> OTHER INFORMATION: common probe; chemically synthetized

<400> SEQUENCE: 49

ctgctgcttt cgagagtggt aacttacg 28

<210> SEQ ID NO 50
<211> LENGTH: 22
<212> TYPE: DNA
<213> ORGANISM: artificial sequence
<220> FEATURE:
<223> OTHER INFORMATION: common probe; chemically synthetized

<400> SEQUENCE: 50

aaacaaactg ctcccggtgt ga 22

<210> SEQ ID NO 51
<211> LENGTH: 36
<212> TYPE: DNA
<213> ORGANISM: artificial sequence
<220> FEATURE:
<223> OTHER INFORMATION: common probe; chemically synthetized

<400> SEQUENCE: 51

attattctac cgatatcagt caataatcaa actcaa 36

<210> SEQ ID NO 52
<211> LENGTH: 25
<212> TYPE: DNA
<213> ORGANISM: artificial sequence
<220> FEATURE:
<223> OTHER INFORMATION: zip-code sequence; chemically synthetized

<400> SEQUENCE: 52

cagccgcggt actgaatgcg atgct 25

<210> SEQ ID NO 53
<211> LENGTH: 25
<212> TYPE: DNA
<213> ORGANISM: artificial sequence
<220> FEATURE:
<223> OTHER INFORMATION: zip-code sequence; chemically synthetized

<400> SEQUENCE: 53

ccccggatag ctgacgaggc ttacg 25

<210> SEQ ID NO 54
<211> LENGTH: 25
<212> TYPE: DNA
<213> ORGANISM: artificial sequence
<220> FEATURE:
<223> OTHER INFORMATION: zip-code sequence; chemically synthetized

<400> SEQUENCE: 54

tccggacagg ttggggtgcg tttgg 25

<210> SEQ ID NO 55
<211> LENGTH: 25
<212> TYPE: DNA
<213> ORGANISM: artificial sequence
<220> FEATURE:
<223> OTHER INFORMATION: zip-code sequence; chemically synthetized

<400> SEQUENCE: 55

-continued

cgtagagcaa cgcgataccc ccgac 25

<210> SEQ ID NO 56
<211> LENGTH: 25
<212> TYPE: DNA
<213> ORGANISM: artificial sequence
<220> FEATURE:
<223> OTHER INFORMATION: zip-code sequence; chemically synthetized

<400> SEQUENCE: 56

cggtcgacga gctgccgcgc aagat 25

<210> SEQ ID NO 57
<211> LENGTH: 25
<212> TYPE: DNA
<213> ORGANISM: artificial sequence
<220> FEATURE:
<223> OTHER INFORMATION: zip-code sequence; chemically synthetized

<400> SEQUENCE: 57

agcagcagtg acaatgccac cgccg 25

<210> SEQ ID NO 58
<211> LENGTH: 25
<212> TYPE: DNA
<213> ORGANISM: artificial sequence
<220> FEATURE:
<223> OTHER INFORMATION: C-zip code sequence; chemically synthetized

<400> SEQUENCE: 58

agcatcgcat tcagtaccgc ggctg 25

<210> SEQ ID NO 59
<211> LENGTH: 25
<212> TYPE: DNA
<213> ORGANISM: artificial sequence
<220> FEATURE:
<223> OTHER INFORMATION: C-zip code sequence; chemically synthethized

<400> SEQUENCE: 59

cgtaagcctc gtcagctatc cgggg 25

<210> SEQ ID NO 60
<211> LENGTH: 25
<212> TYPE: DNA
<213> ORGANISM: artificial sequence
<220> FEATURE:
<223> OTHER INFORMATION: C-zip code sequence; chemically synthetized

<400> SEQUENCE: 60

ccaaacgcac cccaacctgt ccgga 25

<210> SEQ ID NO 61
<211> LENGTH: 26
<212> TYPE: DNA
<213> ORGANISM: artificial sequence
<220> FEATURE:
<223> OTHER INFORMATION: C-zip code sequence; chemically synthetized

<400> SEQUENCE: 61

gtcgggggta tcgcgttgct ctacgg 26

<210> SEQ ID NO 62

-continued

<211> LENGTH: 25
<212> TYPE: DNA
<213> ORGANISM: artificial sequence
<220> FEATURE:
<223> OTHER INFORMATION: C-zip code sequence; chemically synthetized

<400> SEQUENCE: 62

atcttgcgcg gcagctcgtc gaccg 25

<210> SEQ ID NO 63
<211> LENGTH: 25
<212> TYPE: DNA
<213> ORGANISM: artificial sequence
<220> FEATURE:
<223> OTHER INFORMATION: C-zip code sequence; chemically synthetized

<400> SEQUENCE: 63

cggcggtggc attgtcactg ctgct 25

<210> SEQ ID NO 64
<211> LENGTH: 20
<212> TYPE: DNA
<213> ORGANISM: artificial sequence
<220> FEATURE:
<223> OTHER INFORMATION: primer; chemically synthethized

<400> SEQUENCE: 64

gaaatttgtg tagaaggtgc 20

<210> SEQ ID NO 65
<211> LENGTH: 18
<212> TYPE: DNA
<213> ORGANISM: artificial sequence
<220> FEATURE:
<223> OTHER INFORMATION: primer; chemically synthetized

<400> SEQUENCE: 65

caatctcggt atagcggc 18

<210> SEQ ID NO 66
<211> LENGTH: 18
<212> TYPE: DNA
<213> ORGANISM: artificial sequence
<220> FEATURE:
<223> OTHER INFORMATION: primer; chemically synthetized

<400> SEQUENCE: 66

caatgggagc ataacgag 18

<210> SEQ ID NO 67
<211> LENGTH: 3482
<212> TYPE: PRT
<213> ORGANISM: Anabaena

<400> SEQUENCE: 67

Met Ala Gln Asn Thr Asp Tyr Lys Lys Leu Ile Ala Thr Thr Leu Thr
1 5 10 15

Lys Met Glu Ala Met Gln Ala Arg Ile Thr Glu Leu Glu Thr Arg Gln
20 25 30

Ser Glu Pro Ile Ala Val Val Gly Met Gly Cys Arg Phe Pro Gly Gly
35 40 45

Ile Ser Ser Pro Glu Ala Tyr Trp Asn Phe Cys Gln Ala Gly Leu Asp

-continued

```
    50                  55                  60

Ala Ile Val Glu Val Pro Gln Ser Arg Trp Asp Ile Ser Lys Phe Tyr
65                  70                  75                  80

Ala Pro Glu Pro Thr Pro Gly Lys Met Asn Thr Arg Tyr Gly Gly Phe
                85                  90                  95

Leu Gln Gln Asp Ile Thr Glu Phe Asp Ala Arg Phe Phe Ser Ile Ser
            100                 105                 110

Ser Arg Glu Ala Thr Ser Met Asp Pro Gln His Arg Leu Leu Leu Glu
        115                 120                 125

Val Ala Trp Glu Ala Leu Glu Asn Ala Asn Leu Pro Pro Thr Asn Leu
    130                 135                 140

Ala Gly Asp Arg Val Gly Val Phe Val Gly Ile Thr Ser Val Asp His
145                 150                 155                 160

Ala Met Thr Val Tyr Lys Ser Lys Tyr Asp Glu Ile Asp Ser Phe Phe
                165                 170                 175

Gly Thr Gly Asn Ser Leu Ser Ala Ala Ala Gly Arg Leu Ser Tyr Phe
            180                 185                 190

Leu Asn Leu Arg Gly Pro Cys Met Ser Ile Asp Ala Ala Cys Ala Ser
        195                 200                 205

Ser Leu Val Ala Leu His Gln Ala Ile Arg Ser Leu Arg Asn His Glu
    210                 215                 220

Cys Lys Ile Ala Leu Val Gly Gly Val Asn Leu Ile Leu Asp Pro Ala
225                 230                 235                 240

Ile Thr Ile Asn Leu Cys Gln Ser Gly Met Met Ser Pro Asp Gly Arg
                245                 250                 255

Cys Lys Thr Phe Asp Ala Ala Ala Asn Gly Tyr Val Arg Gly Glu Gly
            260                 265                 270

Cys Gly Val Leu Val Leu Lys Arg Leu Ser Val Ala Glu Lys Asn Gly
        275                 280                 285

Asn Arg Ile Leu Ala Leu Leu Arg Gly Ser Ala Val Asn His Asn Gly
    290                 295                 300

Ala Ala Ala Gly Leu Thr Val Pro Ser Gly Pro Ala Gln Gln Asp Leu
305                 310                 315                 320

Leu Arg Gln Ala Leu Ala Asp Ala Arg Val Lys Pro Gln Glu Val Gly
                325                 330                 335

Tyr Ile Glu Ala His Gly Thr Gly Thr Ser Leu Gly Asp Pro Ile Glu
            340                 345                 350

Met Asn Ala Ile Ala Ala Val Tyr Gly Glu Arg Ser Gln Pro Leu Tyr
        355                 360                 365

Val Gly Ser Val Lys Thr Asn Ile Gly His Leu Glu Ala Ala Ala Gly
    370                 375                 380

Ile Ala Gly Thr Ile Lys Thr Ile Leu Ala Leu Gln His Gly Glu Ile
385                 390                 395                 400

Pro Ser His Leu His Phe Gln Glu Pro Asn Pro Leu Ile Asn Trp Gln
                405                 410                 415

Gly Tyr Pro Ile Lys Ile Pro Ser Gln Ala Ile Pro Trp Ser Asn Asn
            420                 425                 430

Gly Gln Val Arg Ile Ala Gly Val Ser Ser Phe Gly Phe Ser Gly Thr
        435                 440                 445

Asn Ala His Ile Ile Ile Glu Gln Ala Pro Ala Ala Asn Ile Pro Glu
    450                 455                 460
```

-continued

```
Ile Lys Leu Gln Arg Pro Ser His Leu Leu Thr Ile Ser Ala His Ser
465                 470                 475                 480

Glu Thr Gly Leu Lys Glu Leu Ala Leu Arg Phe His Thr Arg Leu Glu
                485                 490                 495

Ser His Pro Glu Met Gly Asp Ile Cys His Ser Ala Ala Ile Gly Arg
            500                 505                 510

Ser Ser Leu Pro Glu Arg Leu Ala Ile Val Ala Asp Thr Leu Thr Glu
        515                 520                 525

Leu Gln Gln Arg Leu Ala Ala Phe Ala Glu Glu Lys Asn Ile Asp His
    530                 535                 540

Gly Ile Phe Tyr Arg Arg Phe Thr Gly Glu Lys Tyr Pro Lys Ile Val
545                 550                 555                 560

Phe Leu Phe Thr Gly Gln Gly Ala Cys Tyr Ala Gly Met Gly Asn Gln
                565                 570                 575

Leu Tyr Gln Thr Gln Pro Thr Phe Arg Gln Tyr Ile Asp Gln Cys Ala
            580                 585                 590

Asp Ile Leu Gly Asn Tyr Leu Glu Phe Pro Leu Gln Gln Ile Leu Phe
        595                 600                 605

Gly Asp Arg Thr Asp Leu Leu Asn Gln Thr Ala Tyr Ala Gln Pro Ala
    610                 615                 620

Ile Phe Ala Leu Glu Tyr Ser Leu Ala Met Leu Trp Gln Ser Trp Gly
625                 630                 635                 640

Ile Lys Pro Ser Leu Leu Ile Gly His Ser Val Gly Glu Tyr Val Ala
                645                 650                 655

Ala Cys Ile Ala Gly Val Phe Ser Leu Glu Ala Gly Leu Ala Leu Ile
            660                 665                 670

Val Lys Arg Gly Gln Leu Met Gln Thr Ala Pro Leu Gly Lys Met Ala
        675                 680                 685

Ser Val Phe Ala Asp Glu Ala Thr Val Ser Ala Leu Ile Gln Asn Tyr
    690                 695                 700

Gly Asn Thr Val Ser Ile Ala Ala Ile Asn His Pro Gln Gln Ile Val
705                 710                 715                 720

Ile Ser Gly Glu Ser Asn Ser Ile Asp Glu Ile Val Ala Asn Cys Lys
                725                 730                 735

Ser Gln Lys Ile Ala Val Gln Leu Leu Ser Val Asn Gly Ala Phe His
            740                 745                 750

Ser Pro Leu Met Glu Ser Ile Leu Asp Asp Phe Glu Ile Ala Ala Arg
        755                 760                 765

Glu Val Ser Tyr His Pro Pro Gln Ile Leu Leu Val Ser Gly Ile Asp
    770                 775                 780

Gly Gln Pro Leu Thr Thr Ala Pro Asp Ala Ser Tyr Trp Arg Gln Gln
785                 790                 795                 800

Ser Arg Gln Ala Val Gln Tyr Phe Gln Ser Leu Ile Thr Ala Leu Asn
                805                 810                 815

Lys Gly Tyr Asn Leu Phe Leu Glu Val Gly Pro Arg Pro Ile Leu Ala
            820                 825                 830

Glu Gln Gly Arg Arg Tyr Asn Asp Asp Ala Ile Trp Leu Ser Ser Leu
        835                 840                 845

Asn Arg Gly Leu Asp Asn Trp Gln Thr Met Leu Ser Ala Leu Ala Gln
    850                 855                 860
```

-continued

```
Leu Tyr Ile Asn Gly Val Asn Phe Asn Ala Glu Lys Phe Asn Lys Asp
865                 870                 875                 880

Tyr Gly Tyr Arg Asn Ile Gln Leu Pro Asn Tyr Pro Phe Gln Arg Lys
                885                 890                 895

Arg Phe Gln Phe Lys Ser Thr Val Leu Ser Gln Ser Ser Leu Thr Lys
            900                 905                 910

Glu Val Pro Leu Glu Arg Glu Leu Met Glu Thr Asn Met Asn Leu Ala
        915                 920                 925

Lys Val Ala Asn Ile Lys Lys Asn Gln Gln Glu Ile Gly Asn Lys Leu
    930                 935                 940

Lys Ser Ile Leu Ala Leu Leu Leu Lys Glu Asp Glu Asn Asp Ile Arg
945                 950                 955                 960

Asp Asp Glu Thr Leu Leu Asn Leu Gly Ala Asp Ser Ile Ile Leu Thr
                965                 970                 975

Asp Phe Val Arg Lys Ile Glu Glu Lys Phe Gly Val Lys Val Lys Ile
            980                 985                 990

Asp Gln Leu Phe Thr Asp Leu Gln  Thr Ile Asp Glu Ile  Ser Ile Tyr
        995                 1000                1005

Leu Ser  Asp Tyr Ile Lys Gln  Lys Pro Ser Asn Thr  Ser Asp Glu
    1010                1015                1020

Thr Ala  Ile Asn Asp Ile Leu  Thr Lys Thr Ser Val  Gln Val Ser
    1025                1030                1035

Asn Ser  Glu Ser Glu Leu Asn  Asn Tyr Leu Trp Val  Ile Ser Gln
    1040                1045                1050

Leu Gln  Pro Ile Ala Val Ala  Tyr Ile Leu Lys Ala  Leu Glu Val
    1055                1060                1065

Leu Gly  Lys Arg Leu Ser Ile  Ala Asp Thr Trp Thr  Thr Glu Asp
    1070                1075                1080

Leu Leu  Gln Thr Leu Pro Ile  Ala Ser Lys Tyr His  Ile Leu Val
    1085                1090                1095

Asn Arg  Tyr Leu Lys Thr Leu  Glu Gln Thr Gly Ile  Ile Gln Asn
    1100                1105                1110

Gln Gly  Asn Val Trp Ile Val  Lys Ser Leu Pro Thr  Pro Phe Ser
    1115                1120                1125

Leu Pro  Glu Ala Ile Glu Asn  Leu Gln Thr Ile Cys  Pro Ala Ala
    1130                1135                1140

Lys Pro  Glu Leu Asp Met Leu  Gln Arg Cys Gly Glu  Asn Leu Ala
    1145                1150                1155

Glu Val  Leu Lys Gly Asn Ile  Asp Pro Leu Glu Leu  Ile Phe Pro
    1160                1165                1170

Ala Gly  Ser Val Val His Ala  Glu Ser Ile Tyr Gly  Asn Ser Pro
    1175                1180                1185

Val Ser  Arg Leu Met Asn Gln  Arg Val Ser Gln Ala  Ile Asn Ser
    1190                1195                1200

Ile Leu  Asn Asn Phe Ser Ser  Ser Asp Arg Pro Tyr  Gln Ile Ile
    1205                1210                1215

Glu Val  Gly Gly Gly Thr Gly  Ala Thr Ser Glu Ala  Ile Val Asn
    1220                1225                1230

Asn Leu  Asn Leu Asn His Thr  Thr Tyr Phe Phe Thr  Glu Leu Ser
    1235                1240                1245

Pro Val  Leu Leu Asn Lys Ala  Arg Gln Lys Phe Lys  Asn Arg His
```

-continued

```
    1250                1255                1260

Lys Phe  Asn Phe Asn Gln Leu  Asp Ile Glu Lys Ser  Pro Val Ser
    1265                1270                1275

Gln Gly  Leu Thr Ala His Ser  Tyr His Ile Val Val  Ala Ala Asn
    1280                1285                1290

Val Leu  His Ser Thr Arg Asn  Ile Thr Glu Thr Leu  Asn Asn Ile
    1295                1300                1305

Arg Glu  Leu Leu Ile Pro Gly  Gly Tyr Leu Val Leu  Leu Glu Thr
    1310                1315                1320

Val Glu  Asn Asn Ser Trp Leu  Asp Leu Thr Phe Gly  Leu Thr Pro
    1325                1330                1335

Gly Trp  Trp Arg Phe Gln Asp  Lys Glu Leu Arg Leu  Asp Thr Pro
    1340                1345                1350

Leu Leu  Ser Gly Glu Ser Trp  Cys Ala Ala Leu Lys  Arg Cys Gly
    1355                1360                1365

Phe Val  Asn Ala Asp Ile Tyr  Ser Gln Gln Asn Asn  Ile Ser Ile
    1370                1375                1380

Tyr Asn  Gly Gln Glu Leu Ile  Ile Ala Ser Thr Ser  Pro Glu Ser
    1385                1390                1395

Ala Ile  Asp His Gln Ser Lys  Thr Val Ala Val Ser  Ile Pro Thr
    1400                1405                1410

Ser Gly  Lys Glu Ala Leu Met  Met Ala Gln Leu Gln  Ser Leu Lys
    1415                1420                1425

Glu Leu  Lys Asp Ile His Glu  Lys Thr Ile Ile Lys  Gln Leu Glu
    1430                1435                1440

Ile Leu  Gln Ser Ala Pro Val  Ala Pro Ser Asn Thr  Pro Glu Val
    1445                1450                1455

Leu Leu  Ile Gln Thr Glu Thr  Ala Pro Thr Pro Lys  Ile Ser Lys
    1460                1465                1470

Thr Glu  Thr Thr Pro Pro Thr  Gln Lys Ile Ser Ser  Pro Asn Leu
    1475                1480                1485

Asn Pro  Leu Ala Leu Lys Leu  Thr Glu Ser Lys Ser  Leu Thr Glu
    1490                1495                1500

Gln Gln  Gln Ala Phe Ile Gln  Lys Leu Glu Ile Val  Tyr Asn Gln
    1505                1510                1515

Arg Thr  Ala Lys Ser Lys Ala  Tyr Ser Gln Asn Ser  Arg Lys Thr
    1520                1525                1530

Met Val  Asp Val Lys Pro Thr  Ile Asp Phe Arg Met  Ala Leu Lys
    1535                1540                1545

Glu Phe  Gln Tyr Pro Ile Val  Ser Glu Ser Ala Gln  Gly Ala Tyr
    1550                1555                1560

Phe Arg  Asp Ile Asp Gly Asn  Asp Tyr Ile Asp Leu  Ala Met Gly
    1565                1570                1575

Phe Gly  Val Asn Phe Phe Gly  His Ser Pro Asp Phe  Val Leu Thr
    1580                1585                1590

Glu Ile  Gln Gln Gln Met Gln  His Gly Ile Gly Leu  Gly Met Gln
    1595                1600                1605

Ser Asn  Ile Ala Ala Glu Thr  Ala Ala Leu Ile Cys  Glu Ile Thr
    1610                1615                1620

Gly Val  Glu Arg Val Ala Phe  Ser Asn Thr Gly Thr  Glu Ala Ile
    1625                1630                1635
```

-continued

```
Met Ala  Ala Val Arg Ile Ala  Arg Ser Arg Thr Lys  Arg Gln Lys
    1640                 1645                 1650

Ile Val  Ile Phe Ala Gly Ser  Tyr His Gly Thr Phe  Asp Gly Ile
    1655                 1660                 1665

Leu Ala  Arg Ala Gly Glu Glu  Ala Gly Thr Ala Glu  Pro Leu Ser
    1670                 1675                 1680

Leu Gly  Thr Pro Ser Gly Met  Val Glu Asp Val Ile  Val Leu Thr
    1685                 1690                 1695

Tyr Gly  Ala Glu Glu Ser Leu  Glu Ile Ile Ala Glu  Gln Ala Asp
    1700                 1705                 1710

Asn Leu  Ala Ala Val Leu Val  Glu Pro Val Gln Ser  Arg Lys Pro
    1715                 1720                 1725

Asp Leu  Gln Pro Lys Glu Phe  Ile Gln Lys Leu Arg  Lys Leu Thr
    1730                 1735                 1740

Gln Gln  Lys Glu Ile Ala Leu  Ile Phe Asp Glu Ile  Ile Thr Gly
    1745                 1750                 1755

Phe Arg  Ile Thr Pro Gly Gly  Ala Gln Glu Trp Phe  Glu Ile Glu
    1760                 1765                 1770

Ala Asp  Ile Val Val Tyr Gly  Lys Ala Ile Gly Gly  Gly Leu Pro
    1775                 1780                 1785

Ile Ser  Met Ile Cys Gly Lys  Ala Asp Phe Leu Asp  Thr Val Asp
    1790                 1795                 1800

Gly Gly  Phe Trp Ser Tyr Gly  Asp Asp Ser His Pro  Gln Thr Glu
    1805                 1810                 1815

Leu Thr  Ala Tyr Gly Gly Thr  Phe Cys Arg His Pro  Leu Ala Leu
    1820                 1825                 1830

Ala Ala  Cys Arg Ala Val Leu  Leu His Leu Arg Glu  Gln Gly Ala
    1835                 1840                 1845

Thr Leu  Gln Glu Thr Val Asn  Gln Leu Thr Asn Arg  Leu Ala Ile
    1850                 1855                 1860

Glu Val  Asn Gln Phe Phe Gln  Glu Thr Gly Ile Pro  Ile Arg Ile
    1865                 1870                 1875

Val His  Phe Gly Ser Leu Phe  Arg Phe Glu Ser Ser  Gly Ala Tyr
    1880                 1885                 1890

Ser Ile  Phe Leu Lys Pro Ile  Glu Leu Pro Leu Phe  Tyr Tyr Leu
    1895                 1900                 1905

Leu Asn  Leu Lys Gly Val Tyr  Thr Trp Glu Lys Arg  Val Cys Phe
    1910                 1915                 1920

Leu Ser  Thr Arg His Thr Asn  Glu Asp Ile Asn Lys  Val Val Ala
    1925                 1930                 1935

Ala Val  Lys Glu Ala Ile Ile  Glu Leu Arg Gln Ala  Gly Phe Phe
    1940                 1945                 1950

Ala Asn  Ala Lys Pro Pro Gln  Thr Lys Lys Arg Glu  Ala Ser Asp
    1955                 1960                 1965

Arg Thr  Glu Asp Glu Asp Ala  Arg Asn Asn Leu Asn  Gln Gln Phe
    1970                 1975                 1980

Pro Thr  Ser Glu Ala Gln Arg  Gln Leu Trp Leu Leu  Ala Glu Leu
    1985                 1990                 1995

Asp Thr  Thr Ala Ser Ala Ser  Tyr Asn Val Thr Thr  Ser Leu Glu
    2000                 2005                 2010
```

-continued

```
Leu Arg  Gly Ala Leu Asp Ile  Ser Ser Leu Gln Gln  Ala Ile Asn
    2015                 2020                 2025

Glu Val  Val Asn Arg His Glu  Ala Leu Arg Thr Lys  Ile Leu Glu
    2030                 2035                 2040

Gln Gly  Glu Leu Gln Glu Val  Ile Ser Ser Val Thr  Ile Asp Leu
    2045                 2050                 2055

Pro Leu  Ile Asn Leu Met Asp  Glu Asp Asn Pro Glu  Ala Thr Ala
    2060                 2065                 2070

Leu Val  Leu Arg Thr Glu Leu  Ser Gln Lys Pro Phe  Asp Leu Gly
    2075                 2080                 2085

Val Ala  Pro Leu Phe Ala Ala  Val Leu Met Arg Leu  Ala Pro Glu
    2090                 2095                 2100

His Tyr  Leu Leu Thr Leu Lys  Thr His His Ile Val  Ala Asp Gly
    2105                 2110                 2115

Trp Ser  Leu Gly Leu Ile Leu  Asn Glu Leu Gly Lys  Leu Tyr Ser
    2120                 2125                 2130

Ala Lys  Ile Gly Val Ala Thr  Glu Ser Leu Ser Pro  Pro Met Gln
    2135                 2140                 2145

Phe Arg  Lys Tyr Leu Ala Leu  Arg Gln Gln Glu Ala  Gln Ser Pro
    2150                 2155                 2160

Gln Met  Gln Ala His Arg Asp  Phe Trp Leu Lys Thr  Tyr Glu Gly
    2165                 2170                 2175

Glu Ile  Pro Ile Phe Glu Leu  Pro Thr Asp Phe Pro  Arg Pro Ala
    2180                 2185                 2190

Val Lys  Thr Tyr Thr Gly Gly  Arg Glu Ser Lys Ile  Ile Ala Pro
    2195                 2200                 2205

Gln Leu  Trp Gln Asn Leu Gln  Thr Val Gly Arg Lys  Asn Gln Ala
    2210                 2215                 2220

Thr Leu  Phe Met Thr Met Phe  Ala Ala Tyr Thr Ala  Phe Leu Arg
    2225                 2230                 2235

Arg Ile  Ser Gly His Asp Asp  Leu Val Ile Gly Ile  Pro Ile Ser
    2240                 2245                 2250

Gly Arg  Gln Val Glu Gly Ser  Glu Lys Leu Val Gly  Phe Cys Ser
    2255                 2260                 2265

Gln Phe  Leu Pro Ile Arg Ile  Gln Thr Asp Val Thr  Ala Ser Phe
    2270                 2275                 2280

Val Thr  His Leu Arg His Thr  Lys Glu Thr Leu Ile  Ala Ala Phe
    2285                 2290                 2295

Lys His  Gln Thr His Ala Leu  Glu Glu Leu Leu Ala  Ala Leu Gln
    2300                 2305                 2310

Leu Gln  Arg Asp Phe Ser Arg  Ser Pro Leu Ile Ser  Val Ser Phe
    2315                 2320                 2325

Asn Leu  Asp Pro Lys Leu Thr  Leu Pro Glu Phe Glu  Gly Leu Asn
    2330                 2335                 2340

Val Ser  Leu Pro Pro Glu Pro  Ile Gly Tyr Thr Pro  Phe Asp Leu
    2345                 2350                 2355

Gly Phe  Asn Phe Ile Glu Val  Asn Asp Ala Leu Ile  Ile Tyr Cys
    2360                 2365                 2370

Asn Tyr  Asn Thr Glu Leu Phe  Lys Pro Glu Thr Ile  Lys Gln Phe
    2375                 2380                 2385

Leu Glu  Ser Phe Glu Ile Leu  Met Gln Gly Val Ile  Lys Asp Ala
```

-continued

```
    2390                2395                2400

Asn Ile  Leu Leu Ser Glu Leu  His Leu Leu Thr Gln  Val Gln Gln
    2405                2410                2415

Glu Glu  Leu Leu Ala Lys Leu  Thr Gly Ser Thr Ile  Glu Leu Pro
    2420                2425                2430

Gln Asn  Ser Thr Ile Ile Asp  Asp Phe Ile Ala Gln  Val Lys Ser
    2435                2440                2445

Thr Pro  Asp Ala Pro Ala Leu  Ile Val Glu Glu Lys  Thr Leu Thr
    2450                2455                2460

Tyr Arg  Glu Leu Asn Glu Lys  Val Asn Arg Leu Thr  Asn Tyr Leu
    2465                2470                2475

Arg Glu  Lys Tyr Asn Leu Gly  Ala Gly Lys Ala Ile  Ala Leu Ala
    2480                2485                2490

Ile Gly  Arg Asn Gln Asn Leu  Ile Ile Ala Ile Leu  Ala Thr Phe
    2495                2500                2505

Lys Thr  Gly Ala Ile Tyr Val  Pro Ile Asp Pro Gln  Tyr Pro Ser
    2510                2515                2520

Ser Arg  Ile Asp Phe Ile Leu  Lys Asp Ser Gly Cys  His Leu Cys
    2525                2530                2535

Leu Thr  Glu Ser Asn Phe Ile  Ser Gln Leu Pro Gln  Glu Ile Glu
    2540                2545                2550

Ala Ile  Cys Leu Asp Lys Ile  Asp Asn Ile Leu Thr  Asp Phe Asp
    2555                2560                2565

Ile Asn  Glu Pro Asn Phe Gln  Pro Asp Thr Asn Gln  Ile Ala Tyr
    2570                2575                2580

Ile Leu  Tyr Thr Ser Gly Ser  Thr Gly Asn Pro Lys  Gly Val Met
    2585                2590                2595

Gly Arg  His Ile Ser Ile Leu  Asn Val Ile Arg Ser  Leu Arg Leu
    2600                2605                2610

Thr Phe  Asn Leu Asn Lys His  Pro Glu Trp Arg Tyr  Ile Phe Thr
    2615                2620                2625

Ala Pro  Val Thr His Asp Pro  Ser Phe Arg Asn Ile  Phe Leu Pro
    2630                2635                2640

Leu Thr  Ile Gly Ala Ala Leu  Tyr Met Tyr Glu Val  Gln His Ile
    2645                2650                2655

Gly His  Leu Val Ser Phe Leu  Gln Glu Asn Lys Ile  Asn Val Leu
    2660                2665                2670

His Thr  Thr Pro Ser Ile Tyr  Arg Glu Ile Leu Ala  Val Leu Ala
    2675                2680                2685

Pro Glu  Glu Thr Ile Pro Thr  Leu Lys Tyr Ile Ser  Cys Gly Gly
    2690                2695                2700

Glu Lys  Leu Asp Arg Glu Thr  Ala Ile Ala Leu Arg  Lys Arg Phe
    2705                2710                2715

Pro Ala  Glu Ile Val Ser Asn  Val Tyr Gly Ser Thr  Glu Thr Cys
    2720                2725                2730

Val Gly  Val Ser Gln Tyr Thr  Ile Asp Asp Asn Leu  Asn Thr Asp
    2735                2740                2745

Val Pro  Leu Gly Gln Val Phe  His Asn Asn Arg Leu  Phe Val Leu
    2750                2755                2760

Asp Glu  Phe Asn His Pro Val  Pro Leu His Val Ile  Gly Glu Ile
    2765                2770                2775
```

-continued

```
Cys Val  Glu Gly Ala Ala Leu  Ala Val Gly Tyr Arg  Asn Leu Pro
    2780                2785                2790

Glu Ile  Thr Arg Glu Lys Phe  Gln Pro Asn Phe Leu  Asn Ser Glu
    2795                2800                2805

Lys Ile  Leu Phe Arg Thr Gly  Asp Leu Gly Lys Gln  Ile Ala Pro
    2810                2815                2820

Gly Val  Ile Glu Phe Ile Gly  Arg Lys Asp Asn Gln  Val Lys Val
    2825                2830                2835

Asn Gly  Tyr Arg Val Asp Pro  Gly Glu Ile Glu Tyr  Gln Ile Ser
    2840                2845                2850

Arg Tyr  Ala Glu Ile Glu Lys  Ala Ile Val Leu Pro  Ile Glu Val
    2855                2860                2865

Asn Asn  Gln Ile Gln Leu Ser  Ala Tyr Cys Gln Thr  Asp Lys Asp
    2870                2875                2880

Ile Lys  Phe Ser Glu Ile Arg  Glu Phe Leu Ala Lys  Tyr Leu Pro
    2885                2890                2895

Val Tyr  Met Ile Pro Ser Ser  Phe Ile Phe Leu Lys  Gln Phe Pro
    2900                2905                2910

Leu Thr  Lys His Gly Lys Phe  Asp Leu Arg Ser Leu  Ile Ala Leu
    2915                2920                2925

Lys Pro  Thr Asp Gln Leu Thr  Gln Val Ser Tyr Thr  Ala Pro Arg
    2930                2935                2940

Asn Thr  Leu Glu Ser Lys Leu  Val His Ile Trp Glu  Lys Ile Leu
    2945                2950                2955

Thr Lys  His Pro Ile Gly Ile  Phe Asp Asn Phe Phe  Glu Ile Gly
    2960                2965                2970

Gly His  Ser Leu Leu Leu Ser  Arg Val Val Thr His  Val His Lys
    2975                2980                2985

Glu Leu  Asn Val Leu Val Lys  Leu Ala Asp Phe Phe  Lys Val Pro
    2990                2995                3000

Thr Ile  Leu Gly Leu Ala Ala  Leu Ile Ser Lys Ala  Gln Ser Asn
    3005                3010                3015

Tyr Gln  Glu Pro Ile Pro Ala  Ile Thr Gln Gln Glu  Ser Tyr Pro
    3020                3025                3030

Met Ser  His Gly Gln Arg Arg  Leu Trp Ala Leu Glu  Phe Leu Asp
    3035                3040                3045

His Asn  His Tyr Ala Tyr Gly  Met Pro Ser Ala Tyr  Gln Phe Asn
    3050                3055                3060

Gly Asp  Leu Asn Ile Ala Ala  Phe Glu Asn Ala Phe  Lys Lys Leu
    3065                3070                3075

Ile Glu  Arg His Glu Ile Leu  Arg Thr Thr Phe Thr  Leu Ile Asn
    3080                3085                3090

Asn Glu  Pro Arg Gln Ile Val  Asn Glu Gln Val Asp  Phe Ala Val
    3095                3100                3105

Asn Gln  Ile Asp Leu Val Asp  Asp Glu Asn Gln Ala  Ala Lys Ile
    3110                3115                3120

Ala Glu  Ala Ile Arg Asn Asn  Ala Lys Thr Thr Phe  Asp Leu Glu
    3125                3130                3135

Ser Gly  Leu Leu Leu Lys Ile  Asn Val Leu Lys Leu  Ser Gln Gln
    3140                3145                3150
```

-continued

```
Ser Asn  Ile Val Leu Phe Asn  Met His His Ile Ile  Ser Asp Gly
    3155                 3160                 3165

Trp Ser  Ala Gly Val Leu Ile  Lys Asp Phe Leu Ala  His Tyr His
    3170                 3175                 3180

Ala Tyr  Gly Lys Glu Asn Val  Glu Leu Pro Pro Pro  Leu Arg Ile
    3185                 3190                 3195

His Tyr  Lys Asp Tyr Thr Ser  Trp Gln Asn Gln Gln  Leu Gln Thr
    3200                 3205                 3210

Pro Lys  Leu Gln Ala Gln Arg  Asp Tyr Trp Leu Pro  Lys Leu Ile
    3215                 3220                 3225

Pro Ala  Pro Ala Pro Leu Asp  Leu Pro Leu Asp Tyr  Thr Arg Pro
    3230                 3235                 3240

Ala Val  Gln Ser Phe Ser Gly  Ser Val Val Ile Trp  Lys Pro Asn
    3245                 3250                 3255

Gln Glu  Phe Ile Lys Asp Phe  Glu Leu Leu Thr Lys  Thr Gln Glu
    3260                 3265                 3270

Ala Ser  Leu Phe Met Gly Leu  Leu Thr Leu Val Lys  Gly Phe Leu
    3275                 3280                 3285

Phe Arg  Tyr Thr Glu Gln Asn  Glu Ile Thr Val Gly  Ser Pro Ile
    3290                 3295                 3300

Ala Gly  Arg Asn His Pro Asp  Leu Glu Glu Gln Ile  Gly Phe Tyr
    3305                 3310                 3315

Val Asn  Thr Leu Val Leu Arg  Asp Gln Ile Thr Val  Asp Asp Ser
    3320                 3325                 3330

Phe Ala  Thr Leu Leu Ala Lys  Val Lys Thr Thr Thr  Ile Glu Ala
    3335                 3340                 3345

Tyr Asp  Asn Gln Glu Tyr Pro  Phe Asp Lys Leu Val  Ser Asp Leu
    3350                 3355                 3360

Asn Phe  Lys Arg Asp Pro Ser  Arg Asn Pro Leu Phe  Asp Val Val
    3365                 3370                 3375

Val Val  Leu Gln Asn Asn Gln  Asn Val Asp Leu Ala  Ile Asp Gly
    3380                 3385                 3390

Ile Ala  Val Asn Thr Leu Glu  Gln Glu Leu Val Thr  Ala Lys Phe
    3395                 3400                 3405

Asp Leu  Glu Phe Ile Phe Val  Asp Glu Ala Glu Leu  Tyr Leu Lys
    3410                 3415                 3420

Leu Ile  Tyr Asn Thr Asp Ile  Phe Ala Asn Glu Arg  Ile Ser Leu
    3425                 3430                 3435

Met Ile  Lys Leu Leu Glu Thr  Leu Leu Glu Glu Val  Val Lys Ser
    3440                 3445                 3450

Pro Asp  Thr Pro Leu Leu His  Leu Cys Asp His Thr  Asp Lys Ala
    3455                 3460                 3465

Cys Gln  Glu Asp Asn Ser Leu  Phe Ala Thr Asn Phe  Asn Phe
    3470                 3475                 3480
```

<210> SEQ ID NO 68
<211> LENGTH: 10449
<212> TYPE: DNA
<213> ORGANISM: Anabaena

<400> SEQUENCE: 68

atggctcaaa atacagatta taaaaaatta atagctacaa ccctcacaaa aatggaggcg 60

-continued

atgcaggctc gcattactga attagaaact agacaaagtg agccgattgc tgttgtggga 120

atgggttgtc gtttccctgg aggaatcagt tctcctgaag cttattggaa cttttgtcaa 180

gctggacttg atgcaattgt agaagttcct caaagccgtt gggatatctc aaaatttttat 240

gctccagagc ctactcctgg caaaatgaac actcgttatg ggggattttt acaacaggat 300

attacagaat ttgatgcccg tttcttctcc atatcttccc gcgaagcaac ttcaatggat 360

cctcaacaca ggttattact tgaggtggcg tgggaagcgt tagaaaacgc caatttacca 420

ccaactaatt tagcaggcga tcgcgtgggt gtgtttgtcg gtatcactag tgttgaccac 480

gcgatgacag tctacaaaag caagtatgat gaaatcgatt ctttttttgg tacaggaaac 540

tccctaagtg cagcagcagg taggttatct tattttctca acctccgcgg accttgtatg 600

tctattgatg cagcctgtgc ttcttcattg gttgcacttc accaagctat tcgcagcttg 660

agaaatcatg agtgcaaaat agccttagta ggtggtgtca atctcatctt agatccggca 720

attacgatta acctttgtca gtcggggatg atgtctcccg atggtcgctg caagactttt 780

gatgcggctg caaatggata tgtgcggggc gaaggatgcg gggttttagt tctcaaacgc 840

ctgtctgtcg ctgaaaaaaa tggcaatcgc attctcgcat tactacgggg gtctgcggtc 900

aatcataacg gtgcagccgc cggtttaaca gttcccagtg gccccgccca acaagattta 960

cttcgtcaag ccttagctga tgctagagtc aaaccccagg aagtcggtta tatagaagca 1020

catggtacgg gtacttcttt aggcgacccc attgagatga acgcgatcgc tgctgtgtat 1080

ggagagcgat cgcaacctct ttacgtcggt tcagttaaga ctaatatcgg acatctcgaa 1140

gctgccgcag ggatagcagg cacaatcaag acaattctcg ccctccagca tggtgagatt 1200

ccgtctcatc tccacttcca agaacccaat cccctgatta attggcaagg atacccgatc 1260

aaaattccta gtcaagccat accctggtcg aacaatggcc aagtccgtat cgctggagtc 1320

agttccttcg gtttttctgg aaccaatgct cacataatta ttgaacaagc acctgctgcc 1380

aacataccag aaattaaact gcaacgtccc agccatctgt tgacgatttc tgctcatagc 1440

gaaacgggtt taaaagaact agcactacgt tttcacaccc gcttagagtc tcatccagag 1500

atgggagata tttgtcatag tgcggcaatt ggtaggtcgt ctttacctga acgtttagcg 1560

atcgttgcag atacattgac agagttgcaa caaagattag cagcttttgc tgaggaaaaa 1620

aatattgatc atgggatttt ttatcgacgt tttacaggag aaaaataccc taaaatagtc 1680

tttcttttca ctggtcaagg agcttgttat gcgggtatgg gtaatcaact ttaccaaacc 1740

caaccgacat ttcgccaata tattgatcaa tgtgcagaca ttttaggaaa ttatttagag 1800

tttccactac aacaaatatt atttggcgat cgcacagatt tactcaatca aactgcatac 1860

gctcaaccag ctatatttgc tctagaatat tcacttgcca tgttgtggca atcttgggga 1920

atcaaaccta gtttactaat tggccatagc gttggtgaat atgttgcagc ttgtattgca 1980

ggagtattta gtttagaagc aggtctagca ttaattgtta aacgtgggca attaatgcaa 2040

actgcacctc taggaaaaat ggcttcagtt tttgcagatg aagctacggt atctgctctc 2100

attcaaaact atggtaatac agtttccatc gcggctatta atcatcctca acaaattgtg 2160

atttctgggg aaagtaatag tattgatgag attgtcgcca actgcaaaag ccaaaaaata 2220

gccgttcaat tactatctgt aaatggggct tttcattctc cattaatgga gtcaatctta 2280

gatgattttg aaattgctgc tcgtgaggtt tcgtatcatc ctcctcaaat tctactagtt 2340

-continued

tcaggaattg atggacaacc tttaacaact gcgcctgatg caagttactg gagacaacaa 2400

agccgtcaag cagtgcagta tttccagagt ctaattacgg cgttaaacaa aggatataac 2460

ctatttttag aagtgggtcc aagaccaata ctagcagaac aaggccgtcg ttataacgat 2520

gacgctatat ggcttagttc tctcaataga ggcttggata actggcagac gatgctatca 2580

gcacttgccc aactctacat taatggagtt aattttaacg ccgaaaaatt caataaagat 2640

tatggctata gaaatatcca attaccaaat tatccttttc aaagaaaacg ttttcaattt 2700

aaatctactg ttttatcaca atcaagtttg accaaagaag tgcctttaga aagagagcta 2760

atggagacta atatgaattt agcaaaagtt gcaaatatca aaaaaaatca gcaagaaatc 2820

ggtaataaat taaagtctat cttggctttg ctacttaaag aagatgaaaa tgacattaga 2880

gatgatgaga cattattaaa tctgggagca gattcgatta ttttgaccga ctttgttcgc 2940

aaaattgaag aaaaatttgg cgtaaaagtc aagatagatc aattatttac tgatttacaa 3000

acaattgacg aaatatctat ttatctatct gattatatta aacaaaaacc atcaaataca 3060

tctgatgaaa cagcaataaa tgatatttta actaaaacct cagttcaagt atctaattcc 3120

gagtcagaat taaataatta tctgtgggtg atttcccaat tacaaccaat tgctgttgct 3180

tatatcttga aagctttaga agtattaggt aaaagactaa gtatagcaga cacctggaca 3240

acagaggatt tattacaaac cttacccata gcctcaaaat atcatatttt agttaatcgt 3300

tatttaaaaa ctttagagca aactggaatt attcaaaatc aaggtaatgt atggatagta 3360

aaaagcctac ctacgccttt ttctttacca gaagctatag agaatctgca aacaatttgt 3420

ccagcagcaa aacctgaatt ggatatgcta caacgttgtg gagaaaattt agcagaagtt 3480

ctcaaaggaa atatagatcc tttagagtta attttcccag caggttcagt tgtccatgca 3540

gaaagtatat atggcaattc tcctgtatct cgtttgatga atcaacgagt atcccaagca 3600

attaactcta ttttaaataa tttttctagt agcgatcgcc cttaccaaat tattgaagtg 3660

ggaggtggta caggtgcgac atctgaagca attgttaata atttaaatct caatcataca 3720

acttattttt tcactgaact ttctcctgtt ctcctcaata aagcacggca aaaatttaaa 3780

aataggcata agtttaactt caatcaatta gatattgaaa aatcaccagt ttctcaaggg 3840

ttaacagcgc attcctatca catagttgtt gctgctaacg tgcttcatag tacccgaaat 3900

attacagaaa cactcaataa tattcgagaa ctactaatac ctggtggtta cttggtatta 3960

ctagaaacag tagaaaataa ttcatggctt gatttaactt ttggactcac acctggatgg 4020

tggcgttttc aagataaaga gttacgttta gatactccat tgctaagtgg agaaagttgg 4080

tgtgctgctt taaaacgttg tggatttgtc aatgcagata tttattctca acaaaataat 4140

atcagcatct acaatggaca agagttaatt attgcttcca cttctcctga gagtgctatt 4200

gatcaccaat caaaaacagt agctgtatca ataccaacct ctggcaaaga agctttgatg 4260

atggctcaat tgcaatcttt gaaagaactc aaagatattc atgaaaaaac tattatcaag 4320

caattagaaa tactgcaatc tgcacctgtt gcaccaagca atacaccaga agttttgttg 4380

attcaaacag aaactgcgcc tactcctaaa ataagcaaaa cagaaaccac acctcctacc 4440

caaaaaatta gttctccaaa tttgaatcca ctagctttaa aactgacaga gagtaaatct 4500

ttaaccgaac aacagcaagc tttcatccaa aaattagaga ttgtttataa tcaaagaaca 4560

gcaaaatcca aggcatattc gcaaaacagc cggaagacaa tggttgatgt gaaacctact 4620

-continued

attgacttcc gtatggcttt aaaagaattt cagtatccca tcgtttcaga atcagcccaa 4680

ggtgcatatt ttcgtgatat cgatggcaat gattacatcg acttagcaat ggggtttgga 4740

gttaactttt ttgggcatag tcctgatttt gtgttgacag aaattcagca acaaatgcaa 4800

cacggaattg gtttaggaat gcagtctaat attgcagcag aaacagctgc tttaatttgt 4860

gaaataacag gtgtagaacg agtcgctttt agtaatactg gtacagaagc gattatggca 4920

gcagttcgta ttgctcgttc tcggacaaaa cgccaaaaaa ttgtgatttt tgctggttct 4980

tatcacggta cttttgatgg cattttagcc cgcgctggcg aggaagcagg aacagctgaa 5040

ccattgagtc ttggtacacc atcaggaatg gttgaagatg ttatagttct tacctatgga 5100

gcggaagaaa gtctagaaat aattgcagag caagctgata atttagcagc agttttagta 5160

gaacctgtgc agagccgtaa accagattta cagcctaaag aatttatcca aaaattacgt 5220

aaattgactc aacaaaaaga aattgcccta atttttgatg agattattac tggatttcgc 5280

attactccag gaggagcgca ggaatggttt gagattgagg cagatatagt agtttatggt 5340

aaggcgatcg gtggtggttt gcctatcagt atgatctgcg gtaaagcaga tttcttagac 5400

acagttgatg gcggtttctg gagttatgga gatgattccc atccgcaaac agaattaact 5460

gcttatggtg gtactttctg tcgtcatccc ttagctttgg ctgcttgtcg ggcagtgctt 5520

ttacatttac gtgaacaagg cgcaacgctt caagaaacag tcaatcaatt aactaatcga 5580

ttagctattg aggttaatca gttctttcaa gaaacaggaa ttcctatccg cattgttcat 5640

tttggctctt tgtttcgctt tgaatcttct ggtgcttata gtattttcct caaacctatc 5700

gaattaccac ttttttacta cttattaaat ctcaaaggtg tgtatacatg ggaaaaacga 5760

gtttgttttc tctcaactcg ccataccaat gaggatataa ataaagtagt agctgctgtc 5820

aaagaagcga ttatagaact gcgacaagca ggtttttttcg caaatgctaa accaccacaa 5880

acaaagaaaa gagaagcaag cgatcgcact gaagatgaag atgctaggaa taatttaaat 5940

cagcaatttc ctaccagtga agcgcaacgt cagctttggc tattggcaga attagataca 6000

accgcttcag catcatataa tgtgactact tccttagaat tgcgtggtgc tttagatatt 6060

tcatctttac aacaagcgat taatgaagta gtcaatcgcc acgaagcgtt gcgaaccaaa 6120

atattagaac aaggcgaact acaggaagta atcagtagtg tcacaattga cttaccgtta 6180

atcaatttga tggatgaaga taacccagaa gcaactgctt tagtattgag aactgagtta 6240

tcgcaaaagc cttttgactt gggtgttgct cctttatttg cggctgtgct gatgcgttta 6300

gcacctgaac attatctgct aactttaaaa actcatcata ttgttgcaga tggttggtca 6360

ctagggctaa ttttgaatga actcggcaaa ctttattcag caaaaattgg tgttgctaca 6420

gaatctttat cgccaccaat gcagtttcgc aaatatttag ctttgcgaca gcaagaagcg 6480

caaagtccac aaatgcaagc acatcgtgat ttttggttaa aaacttatga gggagaaatc 6540

cccatatttg aacttcctac agactttcct cgtcctgctg tcaaaactta cacaggcggt 6600

agagaaagta aaattattgc tcctcaacta tggcaaaatt tacagactgt aggacgcaaa 6660

aatcaagcaa cattatttat gacaatgttt gctgcttaca cagctttttt gcggcgcatt 6720

tccggtcatg atgatttagt aattggtatt cccatttctg gacgacaagt cgaaggaagc 6780

gaaaaattag ttggcttttg ttcgcaattt ttaccgattc gtatccagac agatgtcact 6840

gcttcttttg ttacacatct tcgtcacacc aaagaaacac tcatagccgc ttttaaacac 6900

-continued

caaactcacg ccttagaaga attattagca gccttacaat tacaacgaga tttttcgcgt 6960

tctcccctga tttcagtctc tttcaatctc gaccctaaat taactttacc tgaatttgaa 7020

ggactgaacg tatcgctacc accagagcca attggttaca ctcctttcga tttgggcttt 7080

aatttcatcg aagtcaatga cgcactgatt atctactgca actacaatac agaactgttt 7140

aagccggaaa ctatcaaaca gtttctagaa agttttgaaa ttttgatgca gggtgtaatt 7200

aaagatgcca atattttact gtctgagttg catttattaa cacaagtgca acaggaagaa 7260

ttattggcaa aattaacagg ttcaacaata gaattacccc aaaattctac aatcattgat 7320

gactttattg ctcaagttaa atctacaccc gatgcacctg cattaatagt tgaggaaaag 7380

actcttacct atcgagaatt gaatgagaag gtaaatcgtt taaccaatta cttacgtgag 7440

aaatataacc ttggtgcggg gaaggcgatc gcactcgcaa tcggacgcaa tcaaaactta 7500

attatcgcca ttctggctac ttttaaaacc ggggcgatat atgtgccgat agacccccaa 7560

tatcctagca gtcgcatcga ttttatttta aaagatagtg gctgtcatct ctgtcttacc 7620

gagagtaatt ttatctccca attaccccaa gaaatcgaag caatctgctt agataaaatc 7680

gacaatattt tgacggattt tgacataaat gagcctaatt tccaaccaga cactaaccaa 7740

attgcttata ttttatacac ttctggctct accggaaacc ctaaaggagt catgggtcgt 7800

cacatttcta ttttgaatgt aattcggagt ttacgactga catttaactt aaataaacat 7860

cctgaatggc gttatatttt cactgcacct gtcacccatg acccatcttt tcgtaatatt 7920

ttcctacctt taaccatagg tgctgcttta tatatgtacg aagtacagca tatagggcat 7980

ttagtttcat ttttacaaga aaacaaaatc aatgtactgc acacaacacc atctatttac 8040

cgagaaattt tagctgtact cgcacccgaa gaaactatcc ccactttaaa atatatctct 8100

tgcgggggag aaaaattaga ccgagaaact gcgatcgctt tacgaaaacg cttccctgct 8160

gaaattgtca gcaatgtcta cggttctaca gaaacttgtg taggagtatc tcaatataca 8220

atagacgaca acttaaatac agacgtacct cttggtcagg ttttccataa caatcgatta 8280

tttgtcttag atgaattcaa tcaccccgta cctctgcacg taatcggtga aatttgcgta 8340

gaaggtgcag cattagcagt tggttatcgt aatcttcctg aaattaccag agaaaaattt 8400

caacccaact ttttaaattc agaaaaaatt ctctttagaa cgggagattt aggtaaacaa 8460

attgctccag gtgtgattga atttataggg cgaaaagata atcaagttaa ggtcaatggc 8520

tatcgtgtag atccaggaga aattgaatac caaattagcc gctatgccga gattgagaaa 8580

gcaattgtct tacctataga ggtaaataac caaattcaat tatctgctta ttgtcaaact 8640

gataaagata taaaattttc tgaaatcaga gaatttttag ctaaatattt gccagtttac 8700

atgattccta gttcctttat cttcttaaag caatttccct taactaaaca tggcaaattt 8760

gacttgcgat cgctcatcgc tctcaagcca acagatcaat taacacaagt ctcttatact 8820

gcaccgcgta atactttaga atcaaagcta gtccatattt gggaaaaaat tctcactaaa 8880

catcccattg gaatttttga taacttttt gaaatcggcg gacactctct gctcctttct 8940

agagtagtca ctcacgtcca taaagaatta aatgtattag ttaaattagc tgatttcttc 9000

aaagttccca caattcttgg attagcagct ttaatatcta aagctcaatc taactatcaa 9060

gaacccatac cagcaataac tcaacaagaa tcttatccca tgtctcatgg acaacgccgt 9120

ctctgggctt tagaattcct cgatcataac cattatgctt acggaatgcc aagtgcttat 9180

-continued

caattcaatg gtgatttaaa tatcgctgcc tttgaaaatg cttttaaaaa attgatagaa 9240

cgccatgaaa ttttacgcac tacgtttact ttaatcaata acgaacctcg tcaaatagta 9300

aacgaacagg tagattttgc agtcaatcaa atagacttag tagatgatga aaaccaagca 9360

gcaaaaattg ctgaagctat tcgtaacaat gccaaaacta cttttgattt agaatccggt 9420

cttctactca agattaatgt actaaagctg agtcaacaga gtaatatagt cctattcaat 9480

atgcaccata tcatttcaga tggttggtca gcaggcgttt taattaaaga ttttcttgca 9540

cattatcatg cctatggaaa agagaatgta gaattacccc cacctctaag aattcattat 9600

aaagactata cttcatggca aaatcagcaa ctccaaacac caaaattgca agcacaaagg 9660

gattactggt tacctaaact gattcctgct ccagcaccgc ttgatttacc tttagattat 9720

actcgaccag ccgtacaaag tttttctggt tctgtggtta tctggaagcc aaatcaagaa 9780

ttcatcaagg attttgagtt attaactaag actcaagaag ctagtttatt tatgggctta 9840

ctaacattag tcaaaggttt tctgtttcgc tataccgagc agaatgaaat tacagtaggc 9900

tctcccattg ctggaagaaa tcatcccgat ttagaagaac aaataggttt ttatgttaat 9960

actttagtct tgcgcgatca aataacagta gacgatagct ttgctacttt actggcaaaa 10020

gtgaaaacaa ccacaataga ggcatatgat aaccaagaat atccttttga taaactagta 10080

tcagatttaa attttaaacg cgatcctagc cgtaaccctt tatttgatgt tgtagttgtt 10140

ctgcaaaata atcaaaatgt agacctagca atagatggaa ttgctgttaa tacccttgag 10200

caggaacttg ttactgctaa atttgatcta gagttcattt ttgtggacga agcggaatta 10260

tatttgaagt taatctacaa cacagatata ttcgctaacg aacgtatttc attgatgata 10320

aagttgttag aaactcttct agaagaagta gtcaaatctc ctgacacacc attattacat 10380

ctgtgcgatc acacagataa agcctgtcaa gaagataaca gtcttttcgc cacaaatttt 10440

aatttttag 10449

<210> SEQ ID NO 69
<211> LENGTH: 3869
<212> TYPE: PRT
<213> ORGANISM: Anabaena

<400> SEQUENCE: 69

Met Glu Leu Glu Ile Ser Asn Asp Lys Ser Val Gln Ala Arg Val Phe
1 5 10 15

Lys Ala Leu Asn Glu Ala Lys Glu Arg Leu Lys Ala Val Glu Ala Gln
20 25 30

Gln Ser Glu Pro Ile Ala Val Val Gly Leu Gly Cys Arg Phe Gly Lys
35 40 45

Asn Ile Ser Thr Pro Glu Ala Phe Trp Ser Phe Leu Lys Ala Gly Gly
50 55 60

Asn Thr Leu Thr Asp Val Pro Ser Asp Arg Trp Asn Val Ser Asp Phe
65 70 75 80

Tyr Asp Ser Asp Arg Ser Lys Pro Gly Lys Ile Tyr Val Ser Gln Gly
85 90 95

Gly Phe Leu Asn Asp Val Ser Met Phe Asp Ala His Phe Phe Gly Ile
100 105 110

Ala Pro Arg Glu Ala Leu His Ile Asp Pro Gln Gln Arg Leu Leu Leu
115 120 125

-continued

```
Glu Val Gly Tyr Glu Ala Leu Glu Asp Ala Gly Met Ala Ala Pro Thr
    130                 135                 140

Ala Arg Ile Gly Lys Thr Gly Val Phe Ile Gly Ile Thr Asn Asn Asp
145                 150                 155                 160

Tyr Ala Arg Leu Ile Ala Pro Gly Gly Asp Tyr Ser Thr Ile Gly Ala
                165                 170                 175

Tyr His Ile Ser Gly Asn His Ile Asn Ala Ala Ala Gly Arg Ile Ser
            180                 185                 190

Tyr Leu Leu Asn Leu Asn Gly Pro Ser Leu Ala Ile Asp Thr Ala Cys
        195                 200                 205

Ser Ser Ser Leu Val Ala Val His Leu Ala Cys Arg Ser Leu Arg Ser
    210                 215                 220

Gln Glu Cys Arg Gln Ala Leu Val Gly Gly Val Asn Leu Ile Leu Thr
225                 230                 235                 240

Pro Glu Val Pro Ile Ala Leu Cys Arg Asn Gln Met Leu Ala Ala Asp
                245                 250                 255

Gly Arg Cys Lys Thr Phe Asp Glu Ser Ala Asp Gly Phe Gly Ile Gly
            260                 265                 270

Glu Gly Cys Gly Val Val Val Leu Lys Arg Leu Ser Glu Ala Leu Ala
        275                 280                 285

Asp Ser Asp Arg Ile Trp Ala Leu Ile Arg Gly Thr Ser Val Asn Asn
    290                 295                 300

Asp Gly Ala Ser Gly Gly Phe Thr Val Pro Asn Gly Pro Met Gln Thr
305                 310                 315                 320

Asp Leu Ile Arg Asp Ala Leu Ala Asp Ala Arg Leu Ser Ala Asn Ala
                325                 330                 335

Ile Asp Tyr Val Glu Ala His Gly Thr Gly Thr Ser Leu Gly Asp Pro
            340                 345                 350

Ile Glu Ile Lys Ala Ile Ala Asp Ala Leu Cys Val Asn Arg Ser Pro
        355                 360                 365

Ser Gln Pro Leu Leu Ile Gly Ser Val Lys Thr Asn Leu Gly His Leu
    370                 375                 380

Ala Ala Ala Ala Gly Ile Ser Gly Leu Ile Lys Thr Ile Leu Ser Ile
385                 390                 395                 400

Tyr His Gly Glu Ile Pro Ala His Leu Asn Phe Gln Lys Pro Asn Pro
                405                 410                 415

His Ile Asn Trp Asp Ser Ile Pro Leu His Val Val Thr Gln Thr Arg
            420                 425                 430

Pro Trp Thr Gln Gly Asn Gly Lys Ile Lys Ala Ala Gly Val Ser Ser
        435                 440                 445

Phe Gly Ala Ser Gly Thr Asn Ala His Ala Ile Leu Ser Glu Pro Pro
    450                 455                 460

Gln Val Gln Ser Ser Thr Lys Thr Thr Phe Pro Val Leu Val Thr Leu
465                 470                 475                 480

Ser Ala Lys Asp Glu Glu Arg Leu Arg Leu Leu Ala Ala Asn Phe Val
                485                 490                 495

Asn Tyr Ile Asp Gln Lys Pro Glu Val Asn Leu Thr Asp Ile Ala Phe
            500                 505                 510

Ser Val Asn Thr Gly Arg Phe His His Lys Gln Arg Leu Thr Leu Leu
        515                 520                 525

Val Ser Asn Thr Ala Ser Phe Lys Glu Gln Leu Arg Ala Phe Val Asn
```

-continued

```
    530                 535                 540

Gly Asn Asp Arg Glu Ala Asn Ile Ile Tyr Gly Glu Val Arg Val Asn
545                 550                 555                 560

Thr Lys Ile Ala Phe Val Cys Asn Asp Ala Arg Thr Ile Phe Pro Ala
                565                 570                 575

Met Gly Lys Glu Leu Phe Val Asn Glu Pro Val Phe Arg Asp Ala Leu
            580                 585                 590

Met Gln Cys Asp Arg Leu Phe Gln Asn Tyr Leu Asn Phe Ser Ile Glu
        595                 600                 605

Asp Leu Leu Asp Asn Asn Gln Asp Ile Asn Asp Pro Leu Tyr Val His
    610                 615                 620

Pro Thr Leu Phe Ala Leu Gln Tyr Ala Leu Cys Glu Leu Trp Lys Ser
625                 630                 635                 640

Trp Gly Ile Arg Pro Ser Ala Ile Leu Gly Asn Gly Leu Gly Glu Tyr
                645                 650                 655

Ile Ala Ala Gln Gln Ala Gly Ile Phe Ser Ile Gly Asp Ala Val Lys
            660                 665                 670

Leu Val Ala Thr Arg Val His Leu Phe Ser Glu Gln Ser Gly Glu Lys
        675                 680                 685

Glu Phe Leu Ala Phe Gln Lys Val Ala Glu Glu Val Ser Tyr Arg Asn
    690                 695                 700

Pro Gln Leu Thr Leu Ile Ser Asn Asn Gly Glu Val Ala Asn Asn Tyr
705                 710                 715                 720

Ile Lys Thr Ala Ala Tyr Trp Leu Gln Gln Leu His His Leu Asp Asn
                725                 730                 735

Ile Glu Lys Gly Ile Phe Thr Leu His Gln Met Gly Tyr Gln Ile Phe
            740                 745                 750

Leu Glu Ile Gly Asn Gln Pro Ile Asn Asn Ile Val His Ser Lys Leu
        755                 760                 765

Pro Ala Lys Glu Ile Leu Cys Leu Cys Ser Leu Ala Lys Asp Cys Ser
    770                 775                 780

Asn Tyr Gln Gln Ile Gln Thr Val Ile Ala Glu Leu Tyr Val Arg Gly
785                 790                 795                 800

Val Asn Ile Asn Trp Gln Glu Phe His Arg Ser Cys Ser Gly Gln Lys
                805                 810                 815

Ile Ser Leu Pro Ser Ser Pro Phe Ile Arg Lys Arg Tyr Trp Val Asp
            820                 825                 830

Ala Ile Lys Ser Asn Arg Ser Gly His Pro Leu Leu Gly Asp Arg Phe
        835                 840                 845

Pro Ser Pro Leu Ser Ser Ile Gln Tyr Arg Ala Ser Ile Ser Gln Asn
    850                 855                 860

Asn Pro Glu Phe Leu Thr Glu His Gln Val Phe Asp Lys Pro Ile Phe
865                 870                 875                 880

Pro Gly Ala Ala Phe Ile Glu Met Ala Leu Ala Ala Ala Gln Ser Gln
                885                 890                 895

Ala Val Ile Leu Glu Asn Ile Glu Phe Gln Lys Ala Leu Leu Leu Gln
            900                 905                 910

Glu Gln Glu Asp Thr Leu Gln Leu Ile Ile Glu Gln Asn Cys Phe Gln
        915                 920                 925

Ile Tyr Gln Gln Ser Asn Asn Asn Trp Asp Val Leu Val Thr Gly Gly
    930                 935                 940
```

-continued

```
Ile Asp Ser Leu Asn Ile Leu Asp Ala Asn Arg Gln Asn Leu Glu Gln
945                 950                 955                 960

Ile Ala Ala Asn Cys Pro Glu Gln Met Glu Ile Asn Ser Phe Tyr Glu
                965                 970                 975

Arg Tyr Gln Lys Ser Gly Val Asn Tyr Gly Ser Ser Phe Gln Ile Ile
            980                 985                 990

Asn Lys Leu Gln Arg Gly Lys Asn  Ser Ala Phe Ala Lys  Ile Lys Leu
        995                 1000                1005

Thr Gln  Ile Gln Cys Leu Glu  Ala Glu Lys Tyr Lys  Phe His Pro
    1010                1015                1020

Val Met  Leu Asp Ala Cys Phe  Gln Ala Ile Ala Ala  Ile Leu Phe
    1025                1030                1035

Gln Gln  Glu Ser Ser Val Thr  Tyr Val Pro Val Arg  Ile Ala Lys
    1040                1045                1050

Phe Gln  Phe Phe Lys Ser Pro  Gly Ala Ser Val Ile  Ser Ala Ala
    1055                1060                1065

Arg Leu  Thr Lys Asn Ser Asn  Asn Tyr Ile Ile Ser  Asn Ile Asp
    1070                1075                1080

Ile Tyr  Ser Glu Gln Gly Glu  Leu Leu Val Ala Ile  Thr Gly Phe
    1085                1090                1095

Glu Leu  Lys Ser Val Gln Ser  Gln Glu Ile Ile Arg  His Gln Ile
    1100                1105                1110

Gln Pro  Gln Ser Tyr Met Glu  Glu Trp Ile Thr Leu  Ala Ser Leu
    1115                1120                1125

Leu Pro  Asp Gly Arg Asp Tyr  Leu Leu Asn Pro Asp  Ala Ile Lys
    1130                1135                1140

Thr Leu  Val Gln Pro Gln Gln  Leu Glu Gln Gln Leu  Ala Glu Lys
    1145                1150                1155

Leu Tyr  Gln Tyr Glu Arg Leu  Leu Glu Glu Met Glu  Thr Leu Ser
    1160                1165                1170

Val Ser  Tyr Ile Trp Glu Gly  Leu Lys Glu Leu Asn  Trp Gln Pro
    1175                1180                1185

Gln Leu  Gly Gln Ile Tyr Pro  Glu Glu Gln Ile Ala  Thr Gln Gly
    1190                1195                1200

Gly Val  Val Asp Phe Tyr Arg  Pro Leu Leu Ser Arg  Cys Leu Ala
    1205                1210                1215

Ile Leu  Ala Glu Glu Gly Ile  Ile Thr Gln Gln Lys  Asp Gly Trp
    1220                1225                1230

Leu Leu  Ala Lys Glu Pro Val  Ile Ser Ser Ser Gln  Leu Pro Ile
    1235                1240                1245

Gln Gln  Leu Arg Arg Glu Phe  Pro Asp Tyr Leu Ala  Glu Ile Asn
    1250                1255                1260

Leu Ile  Glu Arg Cys Gly Ser  Ala Leu Ala Ala Val  Met Arg Arg
    1265                1270                1275

Gln Ile  Glu Pro Leu Glu Leu  Leu Phe Pro Gln Gly  Asp Leu Asn
    1280                1285                1290

Ala Ile  Ala Ser Val Tyr Ser  Asp Ala Ala Gly Ala  Lys Leu Met
    1295                1300                1305

Asn Glu  Leu Val Ala Ala Thr  Ile Lys Thr Val Val  Ala Asn Leu
    1310                1315                1320
```

-continued

```
Pro Thr  Asn Arg Gln Leu Arg  Ile Leu Glu Ile Gly  Gly Gly Thr
    1325                 1330                 1335

Gly Ser  Ser Thr Ala Ala Ile  Leu Pro His Leu Pro  Pro Glu Gln
    1340                 1345                 1350

Ile Glu  Tyr Thr Phe Thr Asp  Ile Ser Ser Ser Phe  Leu Thr Arg
    1355                 1360                 1365

Ala Lys  Glu Asn Phe Ser Asn  Tyr Pro Phe Ile Lys  Tyr Gln Thr
    1370                 1375                 1380

Leu Asp  Ile Glu Lys Ala Pro  Ile Ser Gln Gly Phe  Leu Pro Ser
    1385                 1390                 1395

Tyr Phe  Asp Ile Ile Ile Ala  Ala Asn Val Leu His  Ala Thr Ala
    1400                 1405                 1410

Asp Ile  Asn Glu Thr Leu Asn  Asn Val Arg Ser Leu  Leu Ala Pro
    1415                 1420                 1425

Asn Ala  Ile Leu Ile Leu Leu  Glu Ser Thr Gly Ala  Arg Pro Trp
    1430                 1435                 1440

Val Asp  Leu Thr Phe Gly Leu  Thr Glu Gly Trp Trp  Leu Cys Ser
    1445                 1450                 1455

Gln Asp  Pro His Arg Asn Gly  Tyr Pro Leu Val Asp  Thr Glu His
    1460                 1465                 1470

Trp Gln  Asn Leu Leu Ala Lys  His Gln Phe Thr Glu  Ile Asn Ile
    1475                 1480                 1485

Ile Glu  Pro Thr Asn Pro Lys  Thr Arg Asn Leu Leu  Gln Gln Ser
    1490                 1495                 1500

Val Ile  Ile Ala Lys Ser Ser  Leu Pro Ser Leu Cys  Thr Ser Val
    1505                 1510                 1515

Ser Trp  Arg Glu Ile Ile Phe  Ala Asp Thr Asn Gly  Ile Ala Arg
    1520                 1525                 1530

Ser Leu  Ile Thr Pro Phe Gln  Gln Arg Gly Ile Thr  Cys Ser Leu
    1535                 1540                 1545

Ile Ser  Pro Gln Asp Ile Asn  Pro Asp Asn Pro Asp  Asp Tyr Leu
    1550                 1555                 1560

Ser Leu  Leu Gln Asn Leu Ile  Thr Pro Glu Thr Arg  Glu Ile Ile
    1565                 1570                 1575

Tyr Leu  Trp Ser Leu Gln Glu  Ile Glu Gly Glu Ile  Tyr Gln Ala
    1580                 1585                 1590

Val Glu  Ile His Cys Arg Arg  Phe Leu Phe Leu Leu  Gln Ala Leu
    1595                 1600                 1605

Leu Gln  Gln Glu Asn Pro Pro  Ala Leu Ile Leu Val  Thr Gln Gly
    1610                 1615                 1620

Ser Val  Pro Ala Lys Glu Ile  Thr Thr Leu Thr Ser  Pro Ala Gln
    1625                 1630                 1635

Ser Ser  Leu Leu Gly Met Ala  Leu Ser Leu Val Leu  Glu His Pro
    1640                 1645                 1650

Glu Leu  Asn Phe Arg Ala Ile  Asp Leu Asp Pro His  Ala Gln Asp
    1655                 1660                 1665

Leu Gly  Glu Lys Leu Phe Arg  Glu Ile His Asn Asn  Thr Gln Glu
    1670                 1675                 1680

Asn Arg  Val Ala Leu Arg Gly  Glu Gln Arg Phe Cys  Pro Arg Leu
    1685                 1690                 1695

Val Glu  Arg Lys Leu Ala Asp  Gly Asn Ile Asn Phe  Arg Gln Asp
```

-continued

```
    1700                1705                1710

Gly Phe  Tyr Leu Ile Ser Gly  Gly Thr Gly Gly Leu  Gly Leu Ala
    1715                1720                1725

Thr Ala  Arg Trp Met Ile Glu  His Gly Ala Cys His  Leu Val Leu
    1730                1735                1740

Cys Ser  Arg Ser Gly Ala Lys  Ala Leu Asn Pro Glu  Ile Leu Ala
    1745                1750                1755

Ser Leu  Gln Ser Ile Asn Glu  Asp Ile Gln Ile Lys  Asp Val Asp
    1760                1765                1770

Val Thr  Asp Ala Glu Lys Leu  His Ala Leu Leu Glu  Glu Cys Arg
    1775                1780                1785

Ser Gln  Tyr Pro Leu Arg Gly  Ile Phe His Ile Ala  Gly Thr Leu
    1790                1795                1800

Asp Asp  Thr Thr Leu Leu Arg  Leu Thr Pro Glu Arg  Phe Asn Tyr
    1805                1810                1815

Val Leu  Ala Pro Lys Val Lys  Gly Thr Trp Leu Leu  His Gln Leu
    1820                1825                1830

Thr Leu  Asn Asp Thr Leu Asp  Phe Phe Val Cys Tyr  Thr Ser Ala
    1835                1840                1845

Val Ser  Leu Ile Gly Ser Ala  Gly Gln Ala Asn Ala  Ala Ala Ala
    1850                1855                1860

Asn Ala  Phe Glu Asp Ala Phe  Thr Tyr Tyr Arg His  Ala His Asn
    1865                1870                1875

Leu Pro  Ala Thr Val Ile Asn  Trp Gly Pro Trp Ser  Glu Ile Gly
    1880                1885                1890

Ala Ala  Val Asp Arg Asn Val  Leu Glu Arg Leu Ala  Ala Lys Gly
    1895                1900                1905

Tyr Asp  Ala Ile Ala Pro Asp  Leu Ala Leu Asn Thr  Leu Glu Lys
    1910                1915                1920

Ile Leu  Phe Asn Gln Ile Val  Arg Ala Gly Val Ile  Ala Ile Asp
    1925                1930                1935

Trp Gln  Arg Phe Pro Tyr Ile  Asn Gln Ser Phe Tyr  Gln Asn Phe
    1940                1945                1950

Leu Pro  Gln Ile Lys Pro Lys  Ser Gln Thr Ala Ser  Asn Leu Leu
    1955                1960                1965

Glu Gln  Trp Gln Ile Ile Pro  Val Lys Gln Arg Arg  Asp Leu Leu
    1970                1975                1980

Ile Arg  Gln Ile Ser Leu Arg  Val Cys Thr Val Leu  Gly Leu Ser
    1985                1990                1995

Thr His  Glu Val Ser Pro Gln  Gln Gly Phe Phe Asp  Met Gly Met
    2000                2005                2010

Asp Ser  Leu Thr Ser Thr Glu  Leu Arg Asn Leu Leu  Gln Thr Asp
    2015                2020                2025

Phe Asn  Cys Ser Leu Pro Thr  Thr Ile Ala Phe Arg  Phe Pro Asn
    2030                2035                2040

Val Glu  Thr Leu Ala Asp Tyr  Leu Leu Arg Glu Ile  Leu Val Thr
    2045                2050                2055

Ser Glu  Val Gln Thr Pro Val  Gln Gln Leu Ile Gln  Glu Val Pro
    2060                2065                2070

Gln Ile  Gln Ile Glu Lys Tyr  Thr Pro Glu Lys Pro  Gln Gln Glu
    2075                2080                2085
```

-continued

```
Glu Asp  Pro Ile Val Ile Val  Gly Met Ala Cys Arg  Phe Pro Gly
    2090                2095                2100

Gly Ala  Asn Asp Leu Glu Ser  Phe Trp Gln Leu Leu  Glu Gln Gly
    2105                2110                2115

Lys Asp  Ala Val Arg Glu Ile  Pro Ser Asp Arg Trp  Asp Met Gln
    2120                2125                2130

Ala Trp  Tyr His Pro Asp Pro  Asp Thr Pro Gly Lys  Ile Tyr Ser
    2135                2140                2145

Pro Tyr  Gly Ala Phe Leu Glu  Gln Ile Asp Gln Phe  Asp Ala Glu
    2150                2155                2160

Phe Phe  Gly Ile Val Pro Arg  Glu Ala Val Ala Ile  Asp Pro Gln
    2165                2170                2175

Gln Arg  Leu Leu Leu Glu Thr  Thr Trp Gln Ala Leu  Glu Ser Ala
    2180                2185                2190

Gly Gln  Asn Pro Gln Lys Leu  Arg Asn Thr Gln Thr  Gly Val Phe
    2195                2200                2205

Val Gly  Ala Met Thr Gln Asp  Tyr Ala Gln Leu Ser  Tyr Ala Pro
    2210                2215                2220

Glu Ala  Ile Asn Ala Tyr Thr  Gly Ser Gly Thr Ser  Leu Ser Val
    2225                2230                2235

Ala Ala  Gly Arg Ile Ser Tyr  Val Leu Gly Leu Gln  Gly Pro Ser
    2240                2245                2250

Met Thr  Val Asp Thr Ala Cys  Ser Ser Ser Leu Val  Ala Val His
    2255                2260                2265

Leu Ala  Cys Asn Ala Leu Arg  Asn Gly Glu Cys Asp  Ile Ala Leu
    2270                2275                2280

Ala Gly  Gly Val Asn Ile Ile  Leu Thr Pro Val Ile  Ser Leu Ile
    2285                2290                2295

Glu Ser  Arg Ala His Met Leu  Ala Pro Asp Gly Arg  Cys Lys Thr
    2300                2305                2310

Phe Asp  Ala Ser Ala Asn Gly  Met Val Arg Gly Glu  Gly Cys Gly
    2315                2320                2325

Met Ile  Val Leu Lys Arg Leu  Ser Gln Ala Val Lys  Ser Gly Asp
    2330                2335                2340

His Ile  Leu Ala Lys Val His  Ser Thr Ala Val Asn  His Asp Gly
    2345                2350                2355

Ser Ser  Ser Gly Leu Thr Val  Pro Asn Gly Asp Ala  Gln Glu Lys
    2360                2365                2370

Leu Leu  His Gln Ala Leu Lys  Ala Ala Lys Leu Asn  Pro Glu Gln
    2375                2380                2385

Ile Asp  Phe Ile Glu Ala His  Gly Thr Gly Thr Ala  Leu Gly Asp
    2390                2395                2400

Pro Ile  Glu Leu Glu Ser Met  Ala Ala Val Phe Gly  Lys Arg Leu
    2405                2410                2415

Gln Asn  Arg Pro Leu Ile Ile  Gly Ser Val Lys Thr  Asn Leu Gly
    2420                2425                2430

His Leu  Glu Gly Ala Ala Gly  Ile Ala Gly Leu Ile  Lys Thr Val
    2435                2440                2445

Leu Ala  Leu Gln His His Lys  Ile Pro Pro His Leu  His Phe Gln
    2450                2455                2460
```

-continued

```
Gln Pro  Asn Pro Arg Phe Asp  Trp Ser Ser Gln Ile  Phe Glu Val
    2465                2470                2475

Pro Val  His Gly Lys Asn Trp  His Pro Ser Gln Arg  Glu Arg Ile
    2480                2485                2490

Ala Gly  Val Ser Ser Phe Gly  Phe Ser Gly Thr Asn  Ala His Ile
    2495                2500                2505

Ile Val  Gly Glu Ile Ala Ser  Asn Ser Pro Gln Pro  Ser Glu Gln
    2510                2515                2520

Lys Phe  Tyr Leu Leu Pro Leu  Ser Ala Arg Ser Gln  Lys Ser Leu
    2525                2530                2535

Lys Glu  Leu Ala Lys Asn Tyr  Gln Tyr Ala Leu Asn  Glu Ser Val
    2540                2545                2550

Asn Phe  Ala Asp Thr Cys Phe  Thr Ala Ser Thr Gly  Arg Ala Ile
    2555                2560                2565

Phe Arg  His Arg Leu Cys Val  Leu Ala Asp Ser Asn  Thr Thr Ala
    2570                2575                2580

Glu Lys  Ala Leu Ala Asp Phe  Gln Lys Gly Glu Asp  Ser Asp Asn
    2585                2590                2595

Leu Ile  Thr Pro Ile Thr Ser  Glu Thr Gln Thr Lys  Val Val Phe
    2600                2605                2610

Leu Phe  Ser Gly Gln Gly Ser  Gln Tyr Ser Gly Met  Gly Gln Thr
    2615                2620                2625

Leu Tyr  Asn Gln Glu Pro Val  Phe Lys Asn Thr Leu  Glu Leu Cys
    2630                2635                2640

Asp Asn  Ile Leu Gln Pro Ile  Leu Gly Lys Ser Leu  Leu Gly Leu
    2645                2650                2655

Ile Phe  Gln Leu Gln Asn Ser  Glu Gln Leu Glu Gln  Thr Gln Ile
    2660                2665                2670

Thr Gln  Pro Ala Leu Phe Ser  Leu Glu Tyr Ala Leu  Ala Lys Leu
    2675                2680                2685

Trp Gln  Ser Trp Gly Ile Gln  Pro Ala Ala Leu Leu  Gly His Ser
    2690                2695                2700

Ile Gly  Glu Tyr Val Ala Ala  Cys Leu Ala Gly Val  Phe Ser Leu
    2705                2710                2715

Glu Asp  Ala Leu Gln Leu Val  Val Gln Arg Gly Arg  Leu Met Gly
    2720                2725                2730

Glu Leu  Pro His Asn Gly Ala  Met Ala Ala Ile Tyr  Ala Asp Tyr
    2735                2740                2745

Gln Thr  Val Ala Asp His Leu  Thr Pro Tyr Gly Asn  Gln Val Asn
    2750                2755                2760

Ile Ala  Ala Asp Asn Gly Ala  Ile Asn Val Ile Ser  Gly Leu Ser
    2765                2770                2775

Glu Ile  Val Glu Gln Leu Glu  Lys Ser Phe Met Glu  Gln Gly Tyr
    2780                2785                2790

Lys Thr  Arg Arg Leu Ala Val  Ser His Ala Phe His  Ser Pro Leu
    2795                2800                2805

Met Glu  Pro Ile Leu Asp Asp  Phe Ala Lys Met Leu  Gln Gln Val
    2810                2815                2820

Ser Phe  His Glu Pro Ser Leu  Asn Ile Ile Ser Asn  Val Thr Gly
    2825                2830                2835

Lys Pro  Ile Gly Lys Glu Ile  Ala Thr Ala Asp Tyr  Trp Leu Arg
```

-continued

```
    2840                2845                2850

His Leu  Arg Asn Thr Val His  Phe Gly Gln Gly Phe  Lys Leu Leu
    2855                2860                2865

Ile Asp  Ser Gly Tyr Arg His  Phe Leu Glu Leu Gly  Pro Lys Pro
    2870                2875                2880

Val Leu  Leu Gly Met Ala Arg  Leu Asn Ser Gln Ser  Lys Glu Ile
    2885                2890                2895

Leu Trp  Leu Pro Ser Ile Val  Pro Gly Gln Asp Glu  Gln Ala Gln
    2900                2905                2910

Met Tyr  Arg Ser Leu Ala Thr  Leu Phe Val Asn Gly  Tyr Ser Ile
    2915                2920                2925

Glu Trp  Thr Glu Val Phe Lys  Gln Val Gln Arg Val  Leu Leu Pro
    2930                2935                2940

Thr Tyr  Pro Phe Gln Arg Glu  Arg Tyr Trp Leu Ser  Asn Pro Gln
    2945                2950                2955

Phe Phe  Val Ser Asp Ile Lys  Thr Lys Leu His Pro  Phe Ile His
    2960                2965                2970

Glu Val  Lys Lys Leu Ala Thr  Gly Glu Ile Ile Val  Glu Gly Glu
    2975                2980                2985

Ile Ser  Ser Glu Tyr Pro Asp  Tyr Leu Glu Gly His  Lys Val Phe
    2990                2995                3000

Gly Lys  Ile Leu Phe Pro Ala  Thr Gly Phe Ile Glu  Thr Ile Leu
    3005                3010                3015

Val Ala  Ser Arg Gln Ile Phe  Thr Asp Asp Leu Ala  Ile Ile Gln
    3020                3025                3030

Asn Val  Ala Ile His Gln Gly  Leu Val Leu Ser Gln  Ser Gln Thr
    3035                3040                3045

Lys Leu  Gln Leu Ile Phe Lys  Ser Lys Lys Ser Ser  Gly Tyr Ser
    3050                3055                3060

Phe Glu  Ile Phe Ser Ser Asp  Ser Ala Lys Asp Ser  Trp Val Leu
    3065                3070                3075

His Val  Thr Gly Glu Ile Asn  Val Asn Glu Arg His  Glu Lys Leu
    3080                3085                3090

Thr His  Arg Arg Arg Asp Ala  Lys Asp Glu Arg Arg  Glu Ser Val
    3095                3100                3105

Asp Leu  Ala Arg Phe Tyr Glu  Phe Tyr Glu Gln Met  Gly Ile Ser
    3110                3115                3120

Tyr Gly  Gln Arg Phe Gln Ala  Ile Gln Glu Leu Arg  Tyr Phe Ser
    3125                3130                3135

Ser Ser  Ser Gln Ala Lys Ile  Ser Ile Asp Arg Ser  Leu Val Asp
    3140                3145                3150

Arg Arg  Tyr Cys Leu His Pro  Val Leu Leu Asp Ala  Cys Leu Gln
    3155                3160                3165

Ser Ile  Gly Ala Ala Phe Pro  Glu Ile His Gly Gln  Glu Leu His
    3170                3175                3180

Leu Pro  Tyr Gly Phe Ser Ser  Leu Glu Leu Phe Gly  Asn Pro Gly
    3185                3190                3195

Thr Gln  Ala Trp Thr His Ile  Glu Ile Lys Ser Gln  Ile Asp Gly
    3200                3205                3210

Glu Ile  Cys Val Asp Val Glu  Val Tyr Asp Glu Gln  Glu Gln Leu
    3215                3220                3225
```

-continued

```
Cys Ala  Arg Phe Thr Asp Leu  Thr Ala Arg Arg Ile  Asn Pro Ala
    3230                3235                3240

Val Leu  Gln Arg Leu Trp Gln  Glu Thr Glu Asn Asn  Cys Phe Tyr
    3245                3250                3255

Gln Val  Gln Trp Gln Lys Leu  Asp Ser Val Ser Thr  Ile Thr Gly
    3260                3265                3270

Asn Ser  Gln His Ser Trp Leu  Val Phe Val Arg Pro  Ser Thr Ala
    3275                3280                3285

Leu Tyr  Gln Ser Ile Asn Leu  Leu Gln Lys Ala Gly  Glu Arg Val
    3290                3295                3300

Ile Thr  Val Glu Leu Ser Asp  Asp Tyr Lys Arg His  Ser Leu Glu
    3305                3310                3315

Ser Phe  Val Ile Asn Pro Ser  Arg Lys Ser Asp Phe  Gln Arg Leu
    3320                3325                3330

Tyr Gln  Glu Ala Tyr Pro Ser  Gly Glu Phe Pro Thr  Gly Val Ile
    3335                3340                3345

Phe Ala  Trp Glu Thr Val Pro  Asn Glu Ala Ser Ala  Asp Thr Val
    3350                3355                3360

Tyr Lys  Ser Cys Asn Ala Val  Leu Tyr Leu Ile Gln  Thr Ile Thr
    3365                3370                3375

Ser Asn  Met Lys Lys Leu Pro  Asp Leu Trp Leu Val  Thr Arg Gly
    3380                3385                3390

Ala Asn  Arg Val Leu Ser Glu  Thr Tyr Leu Gln Pro  Glu Gln Ser
    3395                3400                3405

Pro Leu  Trp Gly Leu Gly Ala  Val Ile Asn His Glu  Tyr Pro Gln
    3410                3415                3420

Ile Arg  Cys Val Cys Leu Asp  Leu Pro Ala Ile Val  Glu Ser His
    3425                3430                3435

Glu Ala  Glu Phe Leu Phe Asn  Glu Phe His Thr Ser  Gly Ser Glu
    3440                3445                3450

Ser Arg  Leu Ala Leu Arg Arg  Gly Asn Arg Tyr Gly  Ala Arg Leu
    3455                3460                3465

Val Ser  Ala Thr Ile Pro Ala  Ala Gln Lys Gln Gln  Leu Val Ser
    3470                3475                3480

Lys Glu  Gly Ala Tyr Leu Ile  Thr Gly Gly Leu Gly  Lys Leu Gly
    3485                3490                3495

Leu Leu  Met Ala Gln Trp Leu  Ser Gln Met Gly Ser  Ser His Leu
    3500                3505                3510

Val Leu  Cys Ser Arg His Val  Lys Ser Gln Pro Glu  Ala Ile Ala
    3515                3520                3525

Ser Leu  Thr Lys Asn Gly Thr  Gln Ile Thr Ile Val  Asn Ala Asp
    3530                3535                3540

Ile Thr  Ser Ala Ala Asp Thr  Glu Gln Leu Phe Ser  Arg Phe Gly
    3545                3550                3555

Ala Asp  Leu Pro Pro Leu Arg  Gly Val Ile His Ala  Ala Ala Val
    3560                3565                3570

Leu Asp  Asp Gly Leu Leu Thr  Asn Gln Asn Trp Glu  Lys Tyr Gln
    3575                3580                3585

Asn Val  Met Arg Pro Lys Val  Glu Gly Thr Leu Leu  Leu Asp Arg
    3590                3595                3600
```

-continued

```
Tyr Thr  Arg Asn Leu Ser Leu  Asp Phe Phe Ile Ala  Phe Ser Ser
    3605                 3610                 3615

Ala Ala  Val Ile Leu Gly Ser  Pro Gly Gln Ser Ser  Tyr Ala Ala
    3620                 3625                 3630

Ala Asn  Ala Phe Met Asp Ala  Leu Ile Gln Gln Arg  Gln Ser Leu
    3635                 3640                 3645

Gly Leu  Pro Gly Ile Ser Ile  Lys Trp Gly Ala Trp  Asp Thr Gly
    3650                 3655                 3660

Asn Lys  Ile Glu Lys Gln Arg  Phe Ala Asn Trp Gly  Ile His Ser
    3665                 3670                 3675

Met Pro  Ser Asp Thr Ala Ile  Lys Tyr Leu Ser Asp  Leu Ile Leu
    3680                 3685                 3690

Ser Asp  Val Asp Gln Gly Ile  Ile Leu Asp Ile Asp  Trp Ser Thr
    3695                 3700                 3705

Phe Asn  Gln Ala Phe Asn Ile  Asn Gln Pro Phe Phe  Ala Glu Val
    3710                 3715                 3720

Ile Thr  Thr Lys Ala Asp Ser  Lys Glu Ala Lys Leu  Leu Glu Arg
    3725                 3730                 3735

Leu Lys  Ser Val Ser Ile Asp  Glu Arg Ala Glu Asn  Leu Ser Gln
    3740                 3745                 3750

Gly Ile  Glu Gln Ile Leu Arg  Glu Val Thr Gly Leu  Ser Ala Ser
    3755                 3760                 3765

Ser Val  Ile Pro His His Thr  Ser Phe Leu Glu Leu  Gly Leu Asn
    3770                 3775                 3780

Ser Leu  Met Val Leu Glu Phe  Lys Asn Arg Leu Gln  Ser Asn Leu
    3785                 3790                 3795

Ala Cys  Thr Leu Pro Thr Ser  Ile Ile Phe Asp Tyr  Pro Asn Ile
    3800                 3805                 3810

Ala Ser  Leu Asn Ile Tyr Leu  Gln Lys Glu Val Leu  Ala Asp Ser
    3815                 3820                 3825

Val Asp  Phe Glu Ile Lys Ser  Asn Glu Ser Ser Glu  Ile Val Asn
    3830                 3835                 3840

Pro Tyr  Glu Ser Leu Asn Glu  Asp Glu Leu Ala Ile  Leu Leu Asn
    3845                 3850                 3855

Gln Lys  Leu Ala Glu Leu Glu  Glu Tyr Gly Asp
    3860                 3865
```

<211> LENGTH: 11610
<212> TYPE: DNA
<213> ORGANISM: Anabaena

<400> SEQUENCE: 70

```
atggagttgg agataagtaa cgataagtct gtccaagcta gggtgtttaa ggcgctaaat     60

gaagcgaagg agagattaaa agctgttgaa gctcaacaaa gcgaaccgat tgcggttgtg    120

ggtttgggtt gtcgttttgg taaaaatatt tccactcctg aagctttttg gtcttttctc    180

aaagctggtg gtaatacgct cacagatgta cctagcgatc gctggaatgt gtctgatttt    240

tatgactcag atcgtagtaa acccggcaaa atctatgtat ctcaaggcgg ttttctcaat    300

gacgtgtcga tgtttgacgc gcacttttttt gggattgcac cgagagaagc tttgcatatc    360

gatcctcagc aacggttact tttagaagtt ggctatgagg cgttagaaga tgctggtatg    420

gctgcgccta ctgccagaat cggcaaaacc ggtgttttta ttggcatcac aaacaatgat    480
```

-continued

tatgctcgtc tgattgcacc gggaggagat tatagcacaa taggggctta tcatatctct 540

ggcaaccata taaatgctgc tgctggacgc atttcttatc ttttgaatct caatggccct 600

agtttggcta tcgatacggc ttgttcttct tcactggtag cagtgcattt agcttgccgt 660

tctttgcgct cccaggaatg tcggcaagct ttggtggggg gagtaaacct aattctcacc 720

ccggaagtgc cgatcgcctt gtgcagaaat cagatgttag ctgctgacgg taggtgcaaa 780

acctttgacg aaagtgctga cggatttggg attggtgaag gttgcggagt tgttgtttta 840

aaaaggttga gtgaagcctt agcagatagt gatcgcattt gggcgttaat tcgaggaaca 900

tctgtcaata acgatggtgc cagtggcggt tttactgttc ctaatggacc tatgcagaca 960

gatttaattc gtgatgcttt agctgatgct cgtctaagtg ccaatgctat tgattatgtc 1020

gaggctcacg gtacgggtac atctttagga gatccgattg aaattaaagc gatcgctgat 1080

gctttatgtg tcaatcgctc tccatcccaa cctttgttaa tcggttctgt caaaacaaat 1140

cttggtcatc tagcggcggc ggcggggatt tctggcttga tcaagactat attatctatc 1200

tatcatggcg aaattccggc acatctcaat ttccaaaaac caaatcctca cattaattgg 1260

gatagcatac ctctgcacgt ggtgacacag actcgcccct ggactcaagg caatggcaag 1320

ataaaagcgg ctggagtcag ttcttttgga gcatcgggaa caaatgctca tgcaatattg 1380

agtgaacctc cgcaagtaca atcctcaact aaaacaacat ttcctgtact ggtaactctg 1440

tctgctaaag atgaagaacg attgcgtctg ttggctgcta actttgttaa ttatattgac 1500

caaaaaccgg aagttaactt aactgatatc gccttttctg taaatacagg tcgttttcat 1560

cataaacagc gattaacttt attagtttcc aatactgcta gttttaaaga acaacttcga 1620

gcttttgtta atggtaatga ccgagaagct aatatcattt acggtgaggt acgagtaaat 1680

actaaaatcg cttttgtttg caatgatgca agaactatat ttccggcgat ggggaaagag 1740

ctttttgtaa atgagcctgt atttcgtgat gctttgatgc aatgcgatcg cctatttcaa 1800

aactatttaa atttctctat agaagactta cttgataaca atcaagatat aaacgatcct 1860

ttatatgttc acccaacatt atttgcgctg caatatgctc tttgtgaact ctggaaatct 1920

tggggaatcc gcccttctgc aattttagga aatggtttag gcgaatatat tgctgctcaa 1980

caagctggaa ttttttcgat tggagatgca gttaaattag tggcaactcg cgttcattta 2040

tttagcgagc aaagtggaga aaaagaattt ttagcttttc aaaaagtagc tgaagaagtt 2100

tcttatcgta atccacaact cactttaatt tctaataatg gtgaggttgc taataattat 2160

attaaaactg ctgcttactg gttgcaacag ctacatcacc tagacaatat tgaaaaaggg 2220

atttttacct tacatcagat gggctatcaa atatttctag agattggaaa tcagcccata 2280

aataatatcg ttcatagtaa attaccagca aaagaaatcc tttgcttatg tagtttagca 2340

aaagattgca gtaattatca gcaaatccaa acagtcatag ctgaactata tgtaagagga 2400

gtaaatatta attggcaaga attccatcgc tcttgctcag gacaaaaaat tagtttgccc 2460

agttctcctt ttattcgcaa acgctactgg gtagatgcta ttaaatcaaa tcgctctgga 2520

cacccttttat taggcgatcg cttcccctct cccctctctt ctattcaata tcgagctagt 2580

attagtcaga acaacccaga atttttaaca gaacatcaag tttttgataa accgatattt 2640

cctggtgcag cttttattga aatggcgtta gcagcggcac aaagccaagc tgtaatttta 2700

gaaaatatcg aatttcaaaa agcactattg ctgcaagaac aagaggatac attacagtta 2760

-continued

attattgaac aaaattgttt tcaaatatat cagcaaagta ataataattg ggatgtactt 2820

gttaccgggg gaattgatag tttaaatata ctagatgcaa atcgccaaaa tctagaacaa 2880

atagctgcta actgtccaga gcagatggaa attaattcat tttatgaaag atatcaaaaa 2940

tcaggagtta attacggcag tagcttccag atcattaata agcttcaacg tggcaaaaat 3000

agtgcatttg ctaaaatcaa gttaacacaa attcagtgtc tagaagccga aaaatataag 3060

tttcatccag taatgttgga tgcttgtttt caggcgatcg ctgcaatttt atttcagcaa 3120

gaatcatctg taacttatgt cccagtccgc attgctaaat ttcaattctt caagtcccca 3180

ggtgctagtg tcatcagtgc agctcgacta actaaaaatt ctaataacta tattatctct 3240

aatattgata tttattcaga acaaggtgag ttactagtag caattacagg atttgaactt 3300

aaatcagttc aaagtcaaga aattatccgc catcaaatcc aaccacaatc ttacatggag 3360

gagtggatta ctttagcatc attgttacct gatggtagag attatttact aaaccctgat 3420

gccattaaaa ctctagttca accccagcaa ttagaacagc aattagctga aaaattgtac 3480

caatatgaac gcttgcttga agaaatggaa actttaagcg taagttatat atgggaaggt 3540

ttaaaagagc ttaactggca accccaatta ggtcaaattt atccagaaga acagattgct 3600

actcaaggtg gagttgttga tttttatcgt cctttattgt cgcggtgttt agcaattcta 3660

gctgaagaag gaataattac tcagcaaaaa gatggatggt tattagcaaa agaacctgtt 3720

atatcttcat ctcaattacc aattcaacaa ctacgtaggg aatttcctga ttaccttgca 3780

gaaattaatt tgatagaacg ttgtggttcg gctttagctg cggtaatgcg ccgccaaatt 3840

gagccattag aattactatt tccccaaggc gatttgaatg cgatcgcatc tgtatattct 3900

gatgcggctg gtgctaagtt gatgaatgaa ctagtcgcag ccacaattaa aactgtagta 3960

gcaaatttac caacaaatcg acagttacgc atcctcgaaa ttggaggggg aacaggctca 4020

agtaccgcag caattttacc acatttacct ccagaacaaa ttgaatacac atttaccgat 4080

atttcttcta gctttttaac cagagccaaa gaaaatttca gtaattatcc cttcattaaa 4140

tatcaaactt tagatataga aaaagccccc atttctcaag gatttctacc tagttatttt 4200

gatattatta ttgctgctaa cgtccttcat gccacagcag atattaatga aacccttaat 4260

aatgtccgtt ctttactagc tcctaatgcc atactcatac tattagaaag taccggagcg 4320

cgtccttggg tagatttaac ctttgggtta acagaaggat ggtggttatg cagtcaagat 4380

ccgcaccgca atggttatcc attagtagat acagaacatt ggcaaaacct actcgccaag 4440

catcaattta cagaaattaa catcatcgaa ccaaccaatc caaagacacg gaacctatta 4500

caacaaagcg taatcatcgc caaaagcagc cttccttctc tatgcacctc tgtgtcttgg 4560

cgcgaaataa tctttgctga cactaacgga attgctcgca gtttaatcac cccattccaa 4620

cagcgaggaa ttacttgttc cctcatctcc cctcaagaca ttaaccccga taaccccgac 4680

gattacctat cactattaca aaacctgata acaccagaaa cccgtgaaat catctacctc 4740

tggtcattac aagaaataga aggcgaaatt tatcaagctg tagaaattca ttgtagacgt 4800

tttctatttc tacttcaggc attattacaa caagaaaacc cccccgcgct cattcttgtt 4860

actcaagggt ctgtaccagc aaaagaaatc acaaccttaa cttctccagc ccaatcttcc 4920

ttattaggaa tggctttaag cttagttcta gagcatcccg aactaaattt tagagcgatt 4980

gatttagatc ctcatgcaca agatttagga gaaaaattat ttagagaaat ccacaacaac 5040

-continued

```
acccaagaaa accgcgtcgc tttgcggggc gaacaacgct tctgtccgcg tttggtagag   5100

cgcaaattag cagatggaaa catcaatttc cgtcaagatg gcttttatct aatttctggt   5160

ggtacaggcg gattaggttt agccactgcc cgatggatga tagaacatgg cgcgtgccat   5220

ttagtattgt gcagtcgcag tggtgcaaaa gctctcaatc ctgaaattct agcaagtctg   5280

caatccatta acgaagacat ccaaattaaa gatgttgatg tcacagatgc cgagaaactg   5340

cacgccctac tagaagaatg ccgatcgcaa tatccgttgc gaggaatttt ccacatcgct   5400

ggcactttag acgatacaac attactgcga ctgacaccag aacgtttcaa ctatgttctc   5460

gcccccaagg tgaaaggaac ttggctactg catcaactaa ccctgaacga taccttagat   5520

ttctttgttt gttatacctc tgctgtgtcg ctcattggtt cggcgggaca agctaatgct   5580

gcggctgcga atgcctttga agatgccttt acctattatc gtcatgcaca caatttacca   5640

gctacagtca tcaactgggg gccgtggagc gaaatcggtg cagcagtcga tcgcaatgtc   5700

ttagaaagat tagcagccaa aggttatgat gcgatcgcac cagatttagc actcaatact   5760

ctagaaaaaa ttctgtttaa tcaaattgta cgtgctggtg taattgctat tgattggcaa   5820

cgttttccat acattaatca aagcttttat caaaacttct taccccaaat caaacctaaa   5880

tctcaaactg catccaatct tttagaacag tggcaaatca tacctgtaaa acagcgtcgt   5940

gatttactta tccgccaaat tagtttgcga gtttgcactg tattaggact atcaactcat   6000

gaagtatctc cccaacaagg ctttttttgac atggggatgg attctctcac ttccacagaa   6060

cttcgcaatc ttctgcaaac agattttaat tgctctctac ccacaacaat cgcttttcgg   6120

tttcccaatg tggaaacttt agcagattat ctcctgcggg aaattttagt tacctctgag   6180

gtgcaaactc ctgtacaaca actaatacaa gaagttcccc aaatccaaat tgaaaaatat   6240

acccccgaaa aaccgcaaca ggaagaagat ccaattgtga ttgtagggat ggcttgtcga   6300

tttcctggcg gtgcgaatga tttggaatca ttctggcaac ttttagaaca ggggaaagat   6360

gcagtcagag aaattcccag cgatcgctgg gatatgcaag catggtatca tcccgatcca   6420

gatacaccag gaaaaatcta ttctccttac ggagcatttc tagagcaaat agatcaattt   6480

gatgccgaat tctttggcat tgtcccccgc gaagcagtcg ccatcgatcc tcagcaacgc   6540

ctactattag aaacaacttg gcaagcctta gaatctgctg ggcaaaaccc ccaaaagtta   6600

cgcaacactc aaacaggggt ttttgtgggg gcgatgactc aagactatgc ccagttgagt   6660

tacgcaccag aagcaattaa cgcctatact ggttctggca cttctctgag tgtagctgcc   6720

ggtcgtattt cgtatgtact gggcttacag ggaccatcaa tgacagttga tactgcttgt   6780

tcttcttctt tagtagcagt acatctagct tgcaatgctt tgcgtaatgg tgagtgtgac   6840

attgccttag caggtggtgt gaatataatt ttaactccgg taatttcctt aattgaaagt   6900

cgcgcccata tgcttgcacc tgatggacgt tgcaagactt ttgatgcgag tgctaatggt   6960

atggtacggg gcgaaggctg cgggatgatt gttctcaagc ggttgagtca agccgttaaa   7020

agtggcgatc acattttagc taaagttcat agtacagccg taaatcacga tggatcaagt   7080

agtggtttga ctgtacctaa tggtgatgct caagaaaaat tattacatca agcattaaag   7140

gcggcgaaac tcaacccaga gcaaatagat tttattgaag ctcatggtac tgggacagct   7200

ttaggcgatc cgattgaatt agaaagtatg gctgcggtat ttggcaaacg tctccaaaat   7260

cgacctttaa ttattggttc tgtcaaaact aatttaggac atttagaagg agcagccgga   7320
```

-continued

attgctgggt taattaaaac tgttttagcc ctacaacatc acaaaattcc tccccatctt 7380

cactttcaac aacccaaccc ccgttttgat tggagttctc agatttttga agttccagta 7440

catggaaaaa actggcatcc tagccaacga gaacgcattg ctggagtaag ttcttttgga 7500

tttagtggta ctaatgctca tattattgtt ggagaaattg catctaattc tccacagcca 7560

tctgagcaga aattttacct cctgccgctt tcggctcgtt ctcaaaaatc tctcaaagaa 7620

ttagcaaaaa attatcaata cgctttaaat gaatctgtga atttcgcaga tacttgtttt 7680

actgccagta caggaagggc tattttccgg catcgattgt gtgtcttggc tgactcaaat 7740

actacagccg aaaaagcact tgctgatttc caaaaaggtg aagattctga taatttaatt 7800

actccaatta catcagaaac tcaaacaaaa gtagttttcc tattttcagg acaaggttct 7860

caatattcag ggatgggaca aactctttac aaccaagaac ccgtctttaa aaatactctg 7920

gaactttgtg acaacattct gcaacctatt ttaggaaagt cgctcttagg tttaattttt 7980

caattgcaaa atagtgaaca gctagaacag actcaaatca cacaaccagc cctgttttca 8040

ctggaatacg cacttgcgaa attatggcaa tcttggggaa tacaacctgc tgccctttta 8100

ggacacagta ttggtgaata tgtcgccgct tgtttagcag gagtcttttc tttagaagat 8160

gctttgcagt tagttgttca acgcggtcgc ttgatgggag aactaccgca taatggagca 8220

atggcagcaa tttatgctga ttatcaaaca gtagccgatc atctgactcc ctatggaaat 8280

caagtaaata ttgccgcaga taatggtgca attaatgtta tttctggctt gagtgaaatt 8340

gttgaacaac tagagaaatc atttatggag caaggttata aaactagacg tttggcagta 8400

tctcatgctt ttcattctcc tttaatggag ccaatccttg atgattttgc taagatgtta 8460

cagcaagttt ctttccatga accaagcctg aatattatta gtaatgtcac aggtaaacca 8520

attggcaaag aaattgctac tgccgattat tggctgcgcc atcttcgcaa tactgtacat 8580

tttggtcaag gatttaaatt gttaattgac agtggttatc gccatttttt ggagttagga 8640

ccaaagcctg tgttattagg tatggcgcgg ctaaattctc aaagcaaaga aatattatgg 8700

ctacccagca ttgtacctgg acaagatgaa caagcccaga tgtacagaag tttagcaaca 8760

ctttttgtta atggttattc tatagaatgg acagaggtat ttaagcaagt ccagcgggtt 8820

cttttaccta cttatccgtt tcaacgagaa cgttattggt tatcaaatcc acaatttttt 8880

gtaagtgata ttaagactaa actgcatcct tttattcatg aggttaaaaa gctagctaca 8940

ggagaaatta ttgttgaagg agagataagt tctgagtatc cagattatct ggaaggacac 9000

aaggtatttg gaaaaatctt gtttccggct acgggtttca ttgaaacgat tttagtagct 9060

agtcggcaaa tatttactga tgatttggcg attattcaaa atgttgctat tcatcagggg 9120

ttagttttat ctcaatctca aactaaattg caactgattt ttaagtctaa gaagtcgtct 9180

ggctatagtt ttgagatatt tagtagtgat tcagcaaaag atagttgggt acttcatgtt 9240

acgggtgaga ttaacgttaa tgaaagacat gaaaagctca cacacagacg cagagacgca 9300

aaggatgaaa gaagggaatc tgtagatcta gccagatttt atgagtttta tgagcagatg 9360

ggaattagct atggtcagcg ttttcaggct attcaagaac tcaggtattt tagcagtagt 9420

tctcaggcta aaattagtat cgatcgcagt ttggtagatc ggcgatattg tttgcatcct 9480

gtgcttttag atgcttgttt acaaagtatt ggcgcggctt tcccagaaat tcacggtcag 9540

gaattgcatt taccttatgg tttctcttct ctagaattat ttggcaatcc tggtactcaa 9600

-continued

gcttggacgc acattgagat taaatctcag attgatggtg aaatctgtgt agatgtagag 9660

gtttatgatg agcaagaaca gctttgcgct cggtttacag acttaacagc acggcggatt 9720

aatccagcag ttttacaacg tttatggcaa gaaactgaaa ataattgttt ttaccaggtg 9780

caatggcaaa agcttgattc agtttcaact attacaggaa actctcaaca ttcatggtta 9840

gtgtttgtac gtcctagcac agctttgtat cagtcgatta atttattgca aaaagctgga 9900

gaaagagtta ttactgtaga acttagtgat gattataaac gtcattctct agaaagtttt 9960

gtcatcaatc catctcggaa atctgatttt caacgtttat atcaagaagc atatccatcg 10020

ggtgaatttc ctacgggtgt gatttttgct tgggaaactg tacctaatga agcaagtgca 10080

gatacagttt acaagagttg taatgctgta ttgtatttaa ttcaaactat tactagtaat 10140

atgaaaaaat tacctgattt atggttagta actcgtggtg cgaatcgagt tttatctgaa 10200

acttatttgc aaccggaaca atctccgctt tggggattag gagcagtaat taatcatgaa 10260

tatccccaaa ttcgttgtgt atgtttagat ttaccagcaa ttgtagaatc tcatgaagct 10320

gaatttttgt tcaacgaatt tcatacatca ggcagtgagt cacgcctagc tttacgccga 10380

ggaaaccgtt atggagcaag gttagtatct gcaacaattc ccgcagccca aaaacagcaa 10440

ttagttagca aagagggagc ttatttaatt acaggtggtt tgggtaaact tggcttgctc 10500

atggctcagt ggttatcaca aatgggatct agccatttgg ttttgtgtag tcgccatgtc 10560

aaatctcaac cagaagcgat cgcctcatta actaaaaatg gtacacagat tacaattgtc 10620

aatgctgata ttacctcagc cgcagatacg gaacagctat tctcccgttt tggcgcggat 10680

ttacccccct tacgcggagt cattcacgcg gcagctgtcc ttgatgatgg gcttttaact 10740

aatcaaaatt gggaaaaata tcaaaatgtt atgcgtccca aagtggaagg cactttatta 10800

ttagaccgct acacccgcaa tctatcttta gacttttca tagccttttc ttctgctgct 10860

gtaattctag gttcaccggg acaaagtagt tacgctgctg ctaatgcttt tatggatgct 10920

ttaatacagc aacgacaaag tttaggatta ccaggtatta gtatcaagtg gggcgcatgg 10980

gatacaggga ataaaattga aaaacaacgt tttgctaatt ggggtattca cagtatgcct 11040

tctgataccg ctatcaaata tctcagtgac ttaattctca gcgacgttga tcaaggaatt 11100

attcttgata ttgattggtc tacttttaat caagctttta acattaatca accattcttt 11160

gccgaagtaa ttacaactaa agcagatagc aaagaggcaa aattattaga acgattaaag 11220

tctgtatcta tagatgagcg agcagaaaat ttatcgcaag gaattgagca aattttacgc 11280

gaagttaccg gattatctgc tagtagtgtg attcctcacc acacctcatt tttagagttg 11340

ggtttaaact cgttaatggt actagaattt aaaaatcggc tgcaaagtaa tttagcttgt 11400

actttaccta catctataat ctttgattat cctaatattg cgagtctcaa catttatttg 11460

cagaaagaag tgttagcaga ttctgttgat tttgagataa aatcaaatga atctagtgag 11520

attgtcaatc cttatgaaag tctcaatgaa gatgagttgg ctatactact taatcaaaag 11580

ttagcggagc ttgaagaata tggtgattaa 11610

<210> SEQ ID NO 71
<211> LENGTH: 24
<212> TYPE: DNA
<213> ORGANISM: artificial sequence
<220> FEATURE:
<223> OTHER INFORMATION: discriminating probe; chemically synthetized

-continued

<400> SEQUENCE: 71

cggtggaaac tggcagacta gaga 24

<210> SEQ ID NO 72
<211> LENGTH: 21
<212> TYPE: DNA
<213> ORGANISM: artificial sequence
<220> FEATURE:
<223> OTHER INFORMATION: discriminating probe; chemically synthetized

<400> SEQUENCE: 72

tgaacactag gtgtcggggg a 21

<210> SEQ ID NO 73
<211> LENGTH: 25
<212> TYPE: DNA
<213> ORGANISM: artificial sequence
<220> FEATURE:
<223> OTHER INFORMATION: discriminating probe; chemically synthethized

<400> SEQUENCE: 73

caagtctgtc gtcaaagaat ggagc 25

<210> SEQ ID NO 74
<211> LENGTH: 22
<212> TYPE: DNA
<213> ORGANISM: artificial sequence
<220> FEATURE:
<223> OTHER INFORMATION: discriminating probe; chemically synthetized

<400> SEQUENCE: 74

acaccatgga agctggcaac at 22

<210> SEQ ID NO 75
<211> LENGTH: 20
<212> TYPE: DNA
<213> ORGANISM: artificial sequence
<220> FEATURE:
<223> OTHER INFORMATION: discriminating probe; chemically synthetized

<400> SEQUENCE: 75

ggctcttggg ctgtcaacca 20

<210> SEQ ID NO 76
<211> LENGTH: 22
<212> TYPE: DNA
<213> ORGANISM: artificial sequence
<220> FEATURE:
<223> OTHER INFORMATION: discriminating probe; chemically synthetized

<400> SEQUENCE: 76

gaaggcgctc tgctgggcca tt 22

<210> SEQ ID NO 77
<211> LENGTH: 24
<212> TYPE: DNA
<213> ORGANISM: artificial sequence
<220> FEATURE:
<223> OTHER INFORMATION: discriminating probe; chemically synthetized

<400> SEQUENCE: 77

cgtaaggaac cagaagacag gtgc 24

-continued

<210> SEQ ID NO 78
<211> LENGTH: 21
<212> TYPE: DNA
<213> ORGANISM: artificial sequence
<220> FEATURE:
<223> OTHER INFORMATION: discriminating probe; chemically synthetized

<400> SEQUENCE: 78

gaaggcgctc tactaggccg c 21

<210> SEQ ID NO 79
<211> LENGTH: 21
<212> TYPE: DNA
<213> ORGANISM: artificial sequence
<220> FEATURE:
<223> OTHER INFORMATION: discriminating probe; chemically synthetized
<220> FEATURE:
<221> NAME/KEY: misc_feature
<222> LOCATION: (15)..(15)
<223> OTHER INFORMATION: N at position 15 is inosine

<400> SEQUENCE: 79

gcgtggggga ggaangctct a 21

<210> SEQ ID NO 80
<211> LENGTH: 25
<212> TYPE: DNA
<213> ORGANISM: artificial sequence
<220> FEATURE:
<223> OTHER INFORMATION: discriminating probe; chemically synthetized

<400> SEQUENCE: 80

cggaatgatt gggcgtaaag ggtct 25

<210> SEQ ID NO 81
<211> LENGTH: 29
<212> TYPE: DNA
<213> ORGANISM: artificial sequence
<220> FEATURE:
<223> OTHER INFORMATION: discriminating probe; chemically synthethized

<400> SEQUENCE: 81

gttaaagagc aaggctcaac cttgtaaag 29

<210> SEQ ID NO 82
<211> LENGTH: 20
<212> TYPE: DNA
<213> ORGANISM: artificial sequence
<220> FEATURE:
<223> OTHER INFORMATION: discriminating probe; chemically synthetized

<400> SEQUENCE: 82

gtcgggacaa cggcagcgag 20

<210> SEQ ID NO 83
<211> LENGTH: 21
<212> TYPE: DNA
<213> ORGANISM: artificial sequence
<220> FEATURE:
<223> OTHER INFORMATION: discriminating probe; chemically synthetized

<400> SEQUENCE: 83

ccggagctca gttcagatcg c 21

-continued

<210> SEQ ID NO 84
<211> LENGTH: 25
<212> TYPE: DNA
<213> ORGANISM: artificial sequence
<220> FEATURE:
<223> OTHER INFORMATION: discriminating probe; chemically synthetized

<400> SEQUENCE: 84

agaaagttgt gaaagcagcc tgacg 25

<210> SEQ ID NO 85
<211> LENGTH: 24
<212> TYPE: DNA
<213> ORGANISM: artificial sequence
<220> FEATURE:
<223> OTHER INFORMATION: discriminating probe; chemically synthetized

<400> SEQUENCE: 85

gaactagagg gcagtagggg taga 24

<210> SEQ ID NO 86
<211> LENGTH: 25
<212> TYPE: DNA
<213> ORGANISM: artificial sequence
<220> FEATURE:
<223> OTHER INFORMATION: discriminating probe; chemically synthetized

<400> SEQUENCE: 86

tgtgccgaag ctaacgcgtt aagtc 25

<210> SEQ ID NO 87
<211> LENGTH: 23
<212> TYPE: DNA
<213> ORGANISM: artificial sequence
<220> FEATURE:
<223> OTHER INFORMATION: discriminating probe; chemically synthetized
<220> FEATURE:
<221> NAME/KEY: misc_feature
<222> LOCATION: (13)..(15)
<223> OTHER INFORMATION: N at position 13 and 15 is inosine

<400> SEQUENCE: 87

gtggcgaaag cgntntgcta gga 23

<210> SEQ ID NO 88
<211> LENGTH: 21
<212> TYPE: DNA
<213> ORGANISM: artificial sequence
<220> FEATURE:
<223> OTHER INFORMATION: discriminating probe; chemically synthetized

<400> SEQUENCE: 88

cgtatcgacc cgtgcagtrc c 21

<210> SEQ ID NO 89
<211> LENGTH: 25
<212> TYPE: DNA
<213> ORGANISM: artificial sequence
<220> FEATURE:
<223> OTHER INFORMATION: discriminating probe; chemically synthethized
<220> FEATURE:
<221> NAME/KEY: misc_feature
<222> LOCATION: (3)..(3)
<223> OTHER INFORMATION: N at position 3 is inosine

-continued

<400> SEQUENCE: 89

ggngtagctc gtatcgaccc gagct 25

<210> SEQ ID NO 90
<211> LENGTH: 23
<212> TYPE: DNA
<213> ORGANISM: artificial sequence
<220> FEATURE:
<223> OTHER INFORMATION: discriminating probe; chemically synthethized

<400> SEQUENCE: 90

gactcctacg ggaggcagca gtg 23

<210> SEQ ID NO 91
<211> LENGTH: 24
<212> TYPE: DNA
<213> ORGANISM: artificial sequence
<220> FEATURE:
<223> OTHER INFORMATION: common probe; chemically synthetized

<400> SEQUENCE: 91

gcagtagggg tagcaggaat tccc 24

<210> SEQ ID NO 92
<211> LENGTH: 22
<212> TYPE: DNA
<213> ORGANISM: artificial sequence
<220> FEATURE:
<223> OTHER INFORMATION: common probe; chemically synthetized
<220> FEATURE:
<221> NAME/KEY: misc_feature
<222> LOCATION: (10)..(10)
<223> OTHER INFORMATION: N at position 10 is inosine

<400> SEQUENCE: 92

atcgaccccn tcggtgtcgt ag 22

<210> SEQ ID NO 93
<211> LENGTH: 29
<212> TYPE: DNA
<213> ORGANISM: artificial sequence
<220> FEATURE:
<223> OTHER INFORMATION: common probe; chemically synthetized
<220> FEATURE:
<221> NAME/KEY: misc_feature
<222> LOCATION: (17)..(17)
<223> OTHER INFORMATION: N at position 17 is inosine

<400> SEQUENCE: 93

ttaactccat aaaggcngtg gaaactgag 29

<210> SEQ ID NO 94
<211> LENGTH: 20
<212> TYPE: DNA
<213> ORGANISM: artificial sequence
<220> FEATURE:
<223> OTHER INFORMATION: common probe; chemically synthetized
<220> FEATURE:
<221> NAME/KEY: misc_feature
<222> LOCATION: (20)..(20)
<223> OTHER INFORMATION: N at position 20 is inosine

<400> SEQUENCE: 94

ccgaagtcgt tactccaacn 20

-continued

<210> SEQ ID NO 95
<211> LENGTH: 27
<212> TYPE: DNA
<213> ORGANISM: artificial sequence
<220> FEATURE:
<223> OTHER INFORMATION: common probe; chemically synthetized
<220> FEATURE:
<221> NAME/KEY: misc_feature
<222> LOCATION: (19)..(19)
<223> OTHER INFORMATION: N at position 19 is inosine

<400> SEQUENCE: 95

cttttctcag ggaagaagnc ctgacgg 27

<210> SEQ ID NO 96
<211> LENGTH: 23
<212> TYPE: DNA
<213> ORGANISM: artificial sequence
<220> FEATURE:
<223> OTHER INFORMATION: common probe; chemically synthetized

<400> SEQUENCE: 96

actgacgctc atggacgaaa gcc 23

<210> SEQ ID NO 97
<211> LENGTH: 22
<212> TYPE: DNA
<213> ORGANISM: artificial sequence
<220> FEATURE:
<223> OTHER INFORMATION: common probe; chemically synthetized

<400> SEQUENCE: 97

tgcatggctg tcgtcagctc gt 22

<210> SEQ ID NO 98
<211> LENGTH: 25
<212> TYPE: DNA
<213> ORGANISM: artificial sequence
<220> FEATURE:
<223> OTHER INFORMATION: common probe; chemically synthetized

<400> SEQUENCE: 98

aactgacact gagggacgaa agcta 25

<210> SEQ ID NO 99
<211> LENGTH: 28
<212> TYPE: DNA
<213> ORGANISM: artificial sequence
<220> FEATURE:
<223> OTHER INFORMATION: common probe; chemically synthetized

<400> SEQUENCE: 99

gggttgtaaa ccccttttct ttgggaag 28

<210> SEQ ID NO 100
<211> LENGTH: 24
<212> TYPE: DNA
<213> ORGANISM: artificial sequence
<220> FEATURE:
<223> OTHER INFORMATION: common probe; chemically synthetized

<400> SEQUENCE: 100

gcaggtggaa ctgaaagtct gctg 24

-continued

<210> SEQ ID NO 101
<211> LENGTH: 28
<212> TYPE: DNA
<213> ORGANISM: artificial sequence
<220> FEATURE:
<223> OTHER INFORMATION: common probe; chemically synthethetized

<400> SEQUENCE: 101

gcagtggaaa ctacatagct agagtgcg 28

<210> SEQ ID NO 102
<211> LENGTH: 23
<212> TYPE: DNA
<213> ORGANISM: artificial sequence
<220> FEATURE:
<223> OTHER INFORMATION: common probe; chemically synthetized

<400> SEQUENCE: 102

ctcgcgagag taagcgaatc cca 23

<210> SEQ ID NO 103
<211> LENGTH: 18
<212> TYPE: DNA
<213> ORGANISM: artificial sequence
<220> FEATURE:
<223> OTHER INFORMATION: common probe; chemically synthetized

<400> SEQUENCE: 103

aggctgcaac tcgcctgc 18

<210> SEQ ID NO 104
<211> LENGTH: 24
<212> TYPE: DNA
<213> ORGANISM: artificial sequence
<220> FEATURE:
<223> OTHER INFORMATION: common probe; chemically synthethized

<400> SEQUENCE: 104

gtaccagagg aatcagcatg gcta 24

<210> SEQ ID NO 105
<211> LENGTH: 22
<212> TYPE: DNA
<213> ORGANISM: artificial sequence
<220> FEATURE:
<223> OTHER INFORMATION: common probe; chemically synthetized

<400> SEQUENCE: 105

gggaattccc ggtgtagcgg tg 22

<210> SEQ ID NO 106
<211> LENGTH: 20
<212> TYPE: DNA
<213> ORGANISM: artificial sequence
<220> FEATURE:
<223> OTHER INFORMATION: common probe; chemically synthetized

<400> SEQUENCE: 106

tcccgcctgg ggagtacgca 20

<210> SEQ ID NO 107
<211> LENGTH: 26

-continued

<212> TYPE: DNA
<213> ORGANISM: artificial sequence
<220> FEATURE:
<223> OTHER INFORMATION: common probe; chemically synthetized

<400> SEQUENCE: 107

caatactgac actgagggac gaaagc 26

<210> SEQ ID NO 108
<211> LENGTH: 25
<212> TYPE: DNA
<213> ORGANISM: artificial sequence
<220> FEATURE:
<223> OTHER INFORMATION: common probe; chemically synthetized
<220> FEATURE:
<221> NAME/KEY: misc_feature
<222> LOCATION: (19)..(19)
<223> OTHER INFORMATION: N at position 19 is inosine

<400> SEQUENCE: 108

gtagctaacg cgttaagtnt cccgc 25

<210> SEQ ID NO 109
<211> LENGTH: 28
<212> TYPE: DNA
<213> ORGANISM: artificial sequence
<220> FEATURE:
<223> OTHER INFORMATION: common probe; chemically synthethized
<220> FEATURE:
<221> NAME/KEY: misc_feature
<222> LOCATION: (7)..(7)
<223> OTHER INFORMATION: N at position 7 is inosine

<400> SEQUENCE: 109

gtrccgnagc taacgcgtta agtatccc 28

<210> SEQ ID NO 110
<211> LENGTH: 20
<212> TYPE: DNA
<213> ORGANISM: artificial sequence
<220> FEATURE:
<223> OTHER INFORMATION: common probe; chemically synthetized

<400> SEQUENCE: 110

gggaattttc cgcaatgggc 20

<210> SEQ ID NO 111
<211> LENGTH: 24
<212> TYPE: DNA
<213> ORGANISM: artificial sequence
<220> FEATURE:
<223> OTHER INFORMATION: C-zip code; chemically synthetized

<400> SEQUENCE: 111

gctgaggtcg atgctgaggt cgca 24

<210> SEQ ID NO 112
<211> LENGTH: 24
<212> TYPE: DNA
<213> ORGANISM: artificial sequence
<220> FEATURE:
<223> OTHER INFORMATION: C-zip code; chemically synthetized

<400> SEQUENCE: 112

gctgcgatcg atggtcaggt gctg 24

-continued

<210> SEQ ID NO 113
<211> LENGTH: 24
<212> TYPE: DNA
<213> ORGANISM: artificial sequence
<220> FEATURE:
<223> OTHER INFORMATION: C-zip code; chemically synthetized

<400> SEQUENCE: 113

gctgtacccg atcgcaaggt ggtc 24

<210> SEQ ID NO 114
<211> LENGTH: 24
<212> TYPE: DNA
<213> ORGANISM: artificial sequence
<220> FEATURE:
<223> OTHER INFORMATION: C-zip code; chemically synthetized

<400> SEQUENCE: 114

cgcaaggtag gtgctgtacc cgca 24

<210> SEQ ID NO 115
<211> LENGTH: 24
<212> TYPE: DNA
<213> ORGANISM: artificial sequence
<220> FEATURE:
<223> OTHER INFORMATION: C-zip code; chemically synthetized

<400> SEQUENCE: 115

cgcacgatag gtggtctacc gctg 24

<210> SEQ ID NO 116
<211> LENGTH: 24
<212> TYPE: DNA
<213> ORGANISM: artificial sequence
<220> FEATURE:
<223> OTHER INFORMATION: C-zip code; chemically synthetized

<400> SEQUENCE: 116

cgcataccag gtcgcatacc ggtc 24

<210> SEQ ID NO 117
<211> LENGTH: 24
<212> TYPE: DNA
<213> ORGANISM: artificial sequence
<220> FEATURE:
<223> OTHER INFORMATION: C-zip code; chemically synthethized

<400> SEQUENCE: 117

ggtcaggtta ccgctgcgat cgca 24

<210> SEQ ID NO 118
<211> LENGTH: 24
<212> TYPE: DNA
<213> ORGANISM: artificial sequence
<220> FEATURE:
<223> OTHER INFORMATION: C-zip code; chemically synthetized

<400> SEQUENCE: 118

ggtccgatta ccggtccgat gctg 24

<210> SEQ ID NO 119
<211> LENGTH: 25

-continued

<212> TYPE: DNA
<213> ORGANISM: artificial sequence
<220> FEATURE:
<223> OTHER INFORMATION: C-zip code; chemically synthetized

<400> SEQUENCE: 119

gggtatccgt tcggtgttgc gtagt 25

<210> SEQ ID NO 120
<211> LENGTH: 25
<212> TYPE: DNA
<213> ORGANISM: artificial sequence
<220> FEATURE:
<223> OTHER INFORMATION: C-zip code; chemically synthetized

<400> SEQUENCE: 120

acctggtcaa tgggaccatt ggtcc 25

<210> SEQ ID NO 121
<211> LENGTH: 25
<212> TYPE: DNA
<213> ORGANISM: artificial sequence
<220> FEATURE:
<223> OTHER INFORMATION: C-zip code; chemically synthetized

<400> SEQUENCE: 121

tatgtcagtg acgcgctcag cgttg 25

<210> SEQ ID NO 122
<211> LENGTH: 25
<212> TYPE: DNA
<213> ORGANISM: artificial sequence
<220> FEATURE:
<223> OTHER INFORMATION: C-zip code; chemically synthetized

<400> SEQUENCE: 122

tggtgctggc gcagaccttt gtctc 25

<210> SEQ ID NO 123
<211> LENGTH: 25
<212> TYPE: DNA
<213> ORGANISM: artificial sequence
<220> FEATURE:
<223> OTHER INFORMATION: C-zip code; chemically synthethized

<400> SEQUENCE: 123

accgcgcaaa tggacagtgt ggcca 25

<210> SEQ ID NO 124
<211> LENGTH: 25
<212> TYPE: DNA
<213> ORGANISM: artificial sequence
<220> FEATURE:
<223> OTHER INFORMATION: C-zip code; chemically synthetized

<400> SEQUENCE: 124

gaccccaact tgacacgtcg caagg 25

<210> SEQ ID NO 125
<211> LENGTH: 25
<212> TYPE: DNA
<213> ORGANISM: artificial sequence
<220> FEATURE:
<223> OTHER INFORMATION: C-zip code; chemically synthetized

-continued

<400> SEQUENCE: 125

ggagagtttg gcgcgaccct aacct 25

<210> SEQ ID NO 126
<211> LENGTH: 25
<212> TYPE: DNA
<213> ORGANISM: artificial sequence
<220> FEATURE:
<223> OTHER INFORMATION: C-zip code; chemically synthetized

<400> SEQUENCE: 126

tgtgcttacc gcacctcgca gtcgt 25

<210> SEQ ID NO 127
<211> LENGTH: 25
<212> TYPE: DNA
<213> ORGANISM: artificial sequence
<220> FEATURE:
<223> OTHER INFORMATION: C-zip code; chemically synthetized

<400> SEQUENCE: 127

gttgggtata tctcccggcg atcgc 25

<210> SEQ ID NO 128
<211> LENGTH: 25
<212> TYPE: DNA
<213> ORGANISM: artificial sequence
<220> FEATURE:
<223> OTHER INFORMATION: C-zip code; chemically synthetized

<400> SEQUENCE: 128

gtattggtgc tcgagtccgg cacga 25

<210> SEQ ID NO 129
<211> LENGTH: 25
<212> TYPE: DNA
<213> ORGANISM: artificial sequence
<220> FEATURE:
<223> OTHER INFORMATION: C-zip code; chemically synthethized

<400> SEQUENCE: 129

gtctacgcca tcgcggtgct aaagc 25

<210> SEQ ID NO 130
<211> LENGTH: 24
<212> TYPE: DNA
<213> ORGANISM: artificial sequence
<220> FEATURE:
<223> OTHER INFORMATION: C-zip code; chemically synthetized

<400> SEQUENCE: 130

ggtctaccta cccgcacgat ggtc 24

<210> SEQ ID NO 131
<211> LENGTH: 784
<212> TYPE: DNA
<213> ORGANISM: Anabaena 90

<400> SEQUENCE: 131

ctccaaaatc gacctttaat tattggttct gtcaaaacta atttaggaca tttagaagga 60

gcagccggaa ttgctgggtt aattaaaact gttttagccc tacaacatca caaaattcct 120

-continued

ccccatcttc actttcaaca acccaacccc cgttttgatt ggagttctca gatttttgaa 180

gttccagtac atggaaaaaa ctggcatcct agccaacgag aacgcattgc tggagtaagt 240

tcttttggat ttagtggtac taatgctcat attattgttg gagaaattgc atctaattct 300

ccacagccat ctgagcagaa attttacctc ctgccgcttt cggctcgttc tcaaaaatct 360

ctcaaagaat tagcaaaaaa ttatcaatac gctttaaatg aatctgtgaa tttcgcagat 420

acttgtttta ctgccagtac aggaagggct attttccggc atcgattgtg tgtcttggct 480

gactcaaata ctacagccga aaaagcactt gctgatttcc aaaaaggtga agattctgat 540

aatttaatta ctccaattac atcagaaact caaacaaaag tagttttcct attttcagga 600

caaggttctc aatattcagg gatgggacaa actctttaca accaagaacc cgtctttaaa 660

aatactctgg aactttgtga caacattctg caacctattt taggaaagtc gctcttaggt 720

ttaatttttc aattgcaaaa tagtgaacag ctagaacaga ctcaaatcac acaaccagcc 780

ctgt 784

<210> SEQ ID NO 132
<211> LENGTH: 784
<212> TYPE: DNA
<213> ORGANISM: Anabaena sp. 202A1/35

<400> SEQUENCE: 132

ctccaaaatc gacctttaat tattggttct gtcaaaacta atttaggaca tttagaagga 60

gcagccggaa ttgctgggtt aattaaaact gttttagccc tacaacatca caaaattcct 120

ccccatcttc actttcaaca acccaacccc cgttttgatt ggagttctca gatttttgaa 180

gttccagtac atggaaaaaa ctggcatcct agccaacgag aacgcattgc tggagtaagt 240

tcttttggat ttagtggtac taatgctcat attattgttg gagaaattgc atctaattct 300

ccacagccat ctgagcagaa attttacctc ctgccgcttt cggctcgttc tcaaaaatct 360

ctcaaagaat tagcaaaaaa ttatcaatac gctttaaatg aatctgtgaa tttcgcagat 420

acttgtttta ctgccagtac aggaagggct attttccggc atcgattgtg tgtcttggct 480

gactcaaata ctacagccga aaaagcactt gctgatttcc aaaaaggtga agattctgat 540

aatttaatta ctccaattac atcagaaact caaacaaaag tagttttcct attttcagga 600

caaggttctc aatattcagg gatgggacaa actctttaca accaagaacc cgtctttaaa 660

aatactctgg aactttgtga caacattctg caacctattt taggaaagtc gctcttagat 720

ttaatttttc aattgcaaaa tagtgaacag ctagaacaga ctcaaatcac acaaccagcc 780

ctgt 784

<210> SEQ ID NO 133
<211> LENGTH: 784
<212> TYPE: DNA
<213> ORGANISM: Anabaena sp. 66A

<400> SEQUENCE: 133

ccccaaaatc aacctttaat tattggttct gtcaaaacta atttaggaca tttagaagga 60

gcagccggaa ttgctgggtt aattaaaact gttttagccc tacaacatca caaaattcct 120

ccccatcttc actttcaaca acccaactcc catttttgatt ggagttctca gatttttgaa 180

gttccagtac atggaaaaaa ctggcatcct agccaacgag aacgcattgc tggagtaagt 240

-continued

tcttttggat ttagtggtcc taatgctcat attattgttg gagaaattgc atctaattct 300

ccacagccat ctgagcagaa attttacctc ctgccgcttt cggctcgttc tcaaaaatct 360

ctcaaagaat tagcaaaaaa ttatcaatac gctttaaatg aatctgtgaa tttcgcagat 420

acttgtttta ctgccagtac aggaagggct gttttccagc atcgattgtg tgtcttagct 480

gactcaaata ctacagccga aaaagcactt gctgatttcc aaaaaggtga agattctgat 540

aatttaatta ctccaattac atcagaaact caaacaaaag tagttttcct attttcagga 600

caaggttctc aatattcagg gatgggacaa actctttaca accaagaacc cgtctttaaa 660

aatactctgg aactttgtga caacattctg caacctattt taggaaagtc gctcttagat 720

ttaatttttc aattgcaaaa tagtgaacag ctagaacaga ctcaaatcac acaaccagcc 780

ctgt 784

<210> SEQ ID NO 134
<211> LENGTH: 784
<212> TYPE: DNA
<213> ORGANISM: Anabaena lemmermannii 202A2

<400> SEQUENCE: 134

ctccaaaatc gacctttaat tattggttct gtcaaaacta atttaggaca tttagaagga 60

gcagccggaa ttgctgggtt aattaaaact gttttagccc tacaacatca caaaattcct 120

ccccatcttc actttcaaca acccaacccc cgttttgatt ggagttctca gatttttgaa 180

gttccagtac atggaaaaaa ctggcatcct agccaacgag aacgcattgc tggagtaagt 240

tcttttggat ttagtggtac taatgctcat attattgttg gagaaattgc atctaattct 300

ccacagccat ctgagcagaa attttacctc ctgccgcttt cggctcgttc tcaaaaatct 360

ctcaaagaat tagcaaaaaa ttatcaatac gctttaaatg aatctgtgaa tttcgcagat 420

acttgtttta ctgccagtac aggaagggct attttccggc atcgattgtg tgtcttggct 480

gactcaaata ctacagccga aaaagcactt gctgatttcc aaaaaggtga agattctgat 540

aatttaatta ctccaattac atcagaaact caaacaaaag tagttttcct attttcagga 600

caaggttctc aatattcagg gatgggacaa actctttaca accaagaacc cgtctttaaa 660

aatactctgg aactttgtga caacattctg caacctattt taggaaagtc gctcttagat 720

ttaatttttc aattgcaaaa tagtgaacag ctagaacaga ctcaaatcac acaaccagcc 780

ctgt 784

<210> SEQ ID NO 135
<211> LENGTH: 784
<212> TYPE: DNA
<213> ORGANISM: Anabaena sp. 299

<400> SEQUENCE: 135

ctccaaaatc gacctttaat tattggttct gtcaaaacta atttaggaca tttagaagga 60

gcagccggaa ttgctgggtt aattaaaact gttttagccc tacaacatca caaaattcct 120

ccccatcttc actttcaaca acccaacccc cgttttgatt ggagttctca gatttttgaa 180

gttccagtac atggaaaaaa ctggcatcct agccaacgag aacgcattgc tggagtaagt 240

tcttttggat ttagtggtac taatgctcat attattgttg gagaaattgc atctaattct 300

ccacagccat ctgagcagaa attttacctc ctgccgcttt cggctcgttc tcaaaaatct 360

-continued

ctcaaagaat tagcaaaaaa ttatcaatac gctttaaatg aatctgtgaa tttcgcagat 420

acttgtttta ctgccagtac aggaagggct attttccggc atcgattgtg tgtcttggct 480

gactcaaata ctacagccga aaaagcactt gctgatttcc aaaaaggtga agattctgat 540

aatttaatta ctccaattac atcagaaact caaacaaaag tagttttcct attttcagga 600

caaggttctc aatattcagg gatgggacaa actctttaca accaagaacc cgtctttaaa 660

aatactctgg aactttgtga caacattctg caacctattt taggaaagtc gctcttaggt 720

ttaatttttc aattgcaaaa tagtgaacag ctagaacaga ctcaaatcac acaaccagcc 780

ctgt 784

<210> SEQ ID NO 136
<211> LENGTH: 784
<212> TYPE: DNA
<213> ORGANISM: Nostoc sp. 152

<400> SEQUENCE: 136

ccccaaaatc gcccttaat tattggttct gtcaaaacta atttagggca tttagaagga 60

gcagccggaa ttgctgggtt aatcaaaact gttttagcct tgcaacatca taaaattcct 120

ccgcatcttc attttgaaaa acccaatccc cgttttgatg ggagttctca tatttttgaa 180

gttccagtac atggtaaaaa ctggcatcct agcgaacgag aaagaattgc gggtgtgagt 240

tccttcggat ttagtggtac gaatgcccat gttattgtgg gagaaattgc atctaatttt 300

tcacaacaat ctgagcatca gctttacctt ttgcctcttt cggctcgttc tgaaaagtcc 360

ctcaaagagt tagccaaaaa ttatcaatct gctttaaatg aatctgtaaa tttagccgat 420

gcttgtttta ccgctagtac aggaagggct aattttcggc atcgattgtg tattctagct 480

gactcaatca ccacagcaga gaaagcgctt actgatttcc agaaaggtga ggattctgag 540

catataatta cgcaaattgc atcagaaact caaccaaata tagctttact attctcagga 600

caaggttctc aatattccgg gatgggacaa actctttaca acaaagaacc tgtctttaaa 660

aatactctag atatttgtga ccaaatcttg caacctattt taggagcatc gctattagat 720

ttaatctttg aagtgtcgaa tagcgatttg ctcgaacaaa ctcaaatcac acagccagcg 780

ctct 784

<210> SEQ ID NO 137
<211> LENGTH: 784
<212> TYPE: DNA
<213> ORGANISM: Nostoc sp. F81

<400> SEQUENCE: 137

ccccaaaatc gacctttaat aattggttcc gtcaaaacca atttaggaca tttagaagga 60

gcagccggaa tagctgggtt aatcaaaact gttttagcct tgcaaaagca tcaaattcct 120

ccccatctcc actttcagca accaaacccc cgttttgatt ggagtaccga tatttttgaa 180

gtcccagtcc atggaaaaaa tgggtatcct agccagcgag aacgcattgc cggcgtgagt 240

tctttgggat ttaggggtac taatgcccat attatggtcg gagaagttgc ccgaaattct 300

ccccaaccat cagagccaaa attttactta ctacccttttt ccgctcgttc cgtaacatct 360

ctgcaacagt tagctaaaaa ttatcagtcc gccttgaatg actctgttaa ttaagcagat 420

gctggtttta ctatcagcac aggtagggct attttccccc atcgattatg tattctcgcc 480

-continued

gactccattt ccacagccca aaaagccctg gctgacttcc aaacaggtga agattcagac 540

aatttaatta cacaaattag ctcagaaact cagcccaaga tcgcctttct cttcactgga 600

cagggttctc aatataccgg catgggagaa acgctttaca accaagaacc agtctttaga 660

aatactctag atatttgtga ccaaatcctc caaccaattt taggaacatc actattagat 720

ttaatctggc aattgccaaa tagcgaatta ctggcacaaa cgcaaatcac acagccagcc 780

ctat 784

<210> SEQ ID NO 138
<211> LENGTH: 784
<212> TYPE: DNA
<213> ORGANISM: Nodularia spumigena HEM

<400> SEQUENCE: 138

ccccaaaatc gacctttaat aattggttcc gtcaaaacca atttaggaca tttagaagga 60

gcagccggaa tagctgggtt aatcaaaact gttttagcct tgcaaaagca tcaaattcct 120

ccccatctcc actttcagca accaaacccc cgtttggatg ggagtaccga tattttggaa 180

gtcccagtac atggaaaaaa tgggtatcct agccagcgag aacgcatggc cggcgtgagt 240

tctttgggat ttaggggtac taatgcccat attatggtcg gagaagttgc ccgaaattct 300

ccccaaccat cagagccaaa attttactta ctacccttt ccgctcgttc cgtaacatct 360

ttgcaacagt tagctaaaaa ttatcagtcc gccttgaatg actctgttaa ttaagcagat 420

gctggtttta ctatcagcac aggtagggct attttccccc atcgattatg tattctcgcc 480

gactccattt ccacagccaa aaaagccctg gctgacttcc aaacaggtga agattcagac 540

aatttaatta cacaaattag ctcagaacct cagcccaaga tcgcctttct cttcactgga 600

caaggttctc aatataccgg catgggagaa acgctttaca accaagaacc agtctttaga 660

aatactctag atatttgtga ccaaatcctc caaccaattt taggaacatc actattagat 720

ttaatctggc aattgccaaa tagcgaatta ctggcacaaa cgcaaatcac acagccagcc 780

ctat 784

<210> SEQ ID NO 139
<211> LENGTH: 784
<212> TYPE: DNA
<213> ORGANISM: Nodularia spumigena BY1

<400> SEQUENCE: 139

ccccaaaatc gacctttaat aattggttcc gtcaaaacca atttaggaca tttagaagga 60

gcagccggaa tagctgggtt aatcaaaact gttttagcct tgcaaaagca taaaattcct 120

ccccatctcc actttcagca accaaacccc cgttttgatg ggagccccga tattttggaa 180

gtcccagtcc atggaaaaaa tgggtatcct agccagcgag aacgcatggc cggcgtgagt 240

tctttgggat ttaggggtac taatgcccat attatggtcg gagaagttgc ccgaatttct 300

ccccaaccat cagagccaaa attttactta ctacccttt ccgttcgttc cgtaacatct 360

ttgcaacagt tagctaaaaa ttatcagtcc gccttgaatg actctgttaa ttaagcagat 420

gctggtttta ctatcagcac aggtagggct attttccccc atcgattatg tattctcgcc 480

gactccattt ccacagccaa aaaagccctg gctgacttcc aaacaggtga agattcagac 540

aatttaatta cacaaattag ctcagaaact cagcccaaga tcgcctttct cttcactgga 600

-continued

caaggttctc aatataccgg catgggagaa acgctttaca accaagaacc agtctttaga 660

aatactctag atatttgtga ccaaatcctc caaccaattt taggaacatc actattagat 720

ttaatctggc aattgccaaa tagcgaatta ctggcacaaa cgcaaatcac acagccagcc 780

ctat 784

<210> SEQ ID NO 140
<211> LENGTH: 784
<212> TYPE: DNA
<213> ORGANISM: Microcystis aeruginosa PCC7941

<400> SEQUENCE: 140

tcccctaatc ggcccttaat tattggttca gtaaaaacga atttaggcca tttagaaggt 60

gctgctggaa ttgccggatt aattaaaacc gttctagcgt tacaacacca taaaatacct 120

cctcatcttc actttaaaaa tcctaatccc cgctttgatt ggagttctca tatttttgaa 180

gttcctgtac aaggaaaacc ttggaatatc agcgaacgtc caaggattgc tggagtaagt 240

tcccttggat ttagtggaac gaatgctcat atcattgttg gggaaattga tgctgattta 300

cctcaagctt cggagaataa tttttatcta ttaccccttt cggctcgttc tgaacaatcc 360

ctgcaagagt tagcgagaag ctatcaagat attttgactg agtctatcaa tttagctgat 420

gtttgtttta ccaccagtac agggcggggg atttttccgc aacgaatctg tattttagca 480

gactcaataa cgacggcaca acgagcatta attgattacc aagatggtga agattctgac 540

tcattaatcc gaccgatttt atcagagact ccgccaaaga tggctttcct cttttctgga 600

caaggctctc aatattctgg catgggagaa actctttata accgagaagt tgtttttaag 660

gaaacattag atctttgtga tcaaatttta gaaccccttt tagaaaaatc tctcctagat 720

ttaatctttc aagagcaaaa tagccagtta ttagaggaaa cccaaatcac tcagccggtg 780

attt 784

<210> SEQ ID NO 141
<211> LENGTH: 783
<212> TYPE: DNA
<213> ORGANISM: Microcystis aeruginosa NIES 89

<400> SEQUENCE: 141

tcccctaatc ggcctttaat tattggttca gtaaaaacga atttaggcca tttagaaggt 60

gctgctggaa ttgccggatt aattaaaacc gttctagcgt tacaacacca taaaatacct 120

cctcatcttc actttaaaaa tcctaatccc cgctttgatt ggagttctca tatttttgaa 180

gttcctgtac aaggaaaacc ttggtatatc agcgaacgtt caaggattgc tggagtaagt 240

tcctttggat ttagtggaac gaatgctcat atcattgttg gggaaattga cgctgattta 300

cctcaacctt cggagaataa tttttatcta ttaccccttt cggctcgttc tgaaaaatcg 360

ctgcaagagt tagcgagaag ttatcaagat attttgactg agtctatcaa tttagctgat 420

gtttgtttta ccgccagtac agggcggggg atttttccgc aacgaatctg tattttagca 480

gactcaatag ccacggcaca acgagcatta attgattacc aagatggtga agattctgac 540

tcattaatcc gaccgatttt atcagagact ccgccaaaga tagctttcct cttttctgga 600

caaggctctc aatattccgg catgggagaa actctttata accgagaagt tgtttttaag 660

gaaacattag atctttgtga tcaaatttta gaaccccttt tagaaaatct ctcctagatt 720

-continued

taatctttca agagcaaaat agccagttat tagaggaaac ccaaatcact cagccggtga 780

ttt 783

<210> SEQ ID NO 142
<211> LENGTH: 784
<212> TYPE: DNA
<213> ORGANISM: Microcystis viridis NIES102

<400> SEQUENCE: 142

tcccctaatc ggcctttaat tattggttca gtaaaaacga atttaggcca tttagaaggt 60

gctgctggaa ttgccggatt aattaaaacc gttctagcgt tacaacacca taaaatacct 120

cctcatcttc actttaaaaa tcctaatccc cgctttgatt ggagttctca tatttttgaa 180

gttcctgtac aaggaaaacc ttgggatatc agcgaacgtt caaggattgc tggagtaagt 240

tcctttggat ttagtggaac gaatgctcat atcattgttg gggaaattga cgctgattta 300

cctcaacctt cggagaataa tttttatcta ttaccccttt cggctcgttc tgaaaaatcg 360

ctgcaagagt tagcgagaag ttatcaagat attttgactg agtctatcaa tttagctgat 420

gtttgtttta ccgccagtac agggcggggg atttttccgc aacgaatctg tattttagca 480

gactcaatag ccacggcaca acgagcatta attgattacc aagatggtga agattctgac 540

tcattaatcc gaccgatttt atcagagact ccgccaaaga tagctttcct cttttctgga 600

caaggctctc aatattccgg catgggagaa acccttttata accgagaagt tgttttcaag 660

gaaacattag atctttgtga tcaaattcta gaaccccttt tagaaaaatc tctcttagat 720

ttaatctttc aagagcaaaa tagtgagtta ttagaggaaa cccaaatcac tcagccggtg 780

attt 784

<210> SEQ ID NO 143
<211> LENGTH: 784
<212> TYPE: DNA
<213> ORGANISM: Microcystis sp. K139

<400> SEQUENCE: 143

tcccctaatc ggcccttaat tattggttca gtaaaaacga atttaggcca tttagaaggt 60

gctgccggaa ttgccggatt aattaaaacc gttctagcgt tacaacacca taaaatacct 120

cctcatcttc actttaaaaa tcctaatccc cgctttgatt ggagttctca tatttttgaa 180

gttcctgtac aaggaaaacc ttgggatatc agcgaacgtc caaggattgc tggagtaagt 240

tcctttggat ttagtggaac gaatgctcat atcattgttg gggaaattga tgctgatttg 300

cctcaagctt cggagaataa tttttatcta ttaccccttt cggctcgttc tgaacaatcc 360

ctgcaagagt tagcgagaag ttatcaagat attttgactg agtctatcaa tttagctgat 420

gtttgtttta ccaccagtac agggcggggg atttttccgc aacgaatctg tattttagca 480

gactcaataa ccacggcaca acgagcatta attgattacc aagatggtga aaattctgac 540

tcattaatcc gaccgatttt atcagagact ccacaaaaga tagctttcct cttttctgga 600

caaggctctc aatattccgg catgggagaa acccttttata accgagaagt tgttttcaag 660

gaaacattag atctttgtga tcaaatttta gaaccccttt tagaaaaatc tctcttagat 720

ttaatctttc aagagcaaaa tagtgagtta ttagaggaaa cccaaatcac tcagccggtg 780

attt 784

-continued

<210> SEQ ID NO 144
<211> LENGTH: 784
<212> TYPE: DNA
<213> ORGANISM: Microcystis aeruginosa PCC7806

<400> SEQUENCE: 144

tcccctaatc ggcccttaat tattggttca gtaaaaacga atttaggcca tttagaaggt 60

gctgccggaa ttgccggatt aattaaaacc gttctagcgt tacaacacca taaaatacct 120

cctcatcttc actttaaaaa tcctaatccc cgctttgatt ggagttctca tatttttgaa 180

gttcctgtac aaggaaaacc ttggaatatc agcgaacgtc caaggattgc tggagtaagt 240

tcctttggat ttagtggaac gaatgctcat atcattgttg gggaaattga tgctgatttg 300

cctcaagctt cggagaataa tttttatcta ttaccccttt cggctcgttc tgaacaatcc 360

ctgcaagagt tagccagaag ctatcaagat attttgactg agtctatcaa tttagctgat 420

gtttgtttta ccaccagtac agggcggggg atttttccgc aacgaatctg tattttagca 480

gactcaataa cgacggcaca acgagcatta attgattacc aagatggtga agattctgac 540

tcattaatcc gaccgatttt atcagagact ccgccaaaga tatctttcct cttttctgga 600

caaggctctc aatattctgg catgggagaa actctttata accgagaagt tgtttttaag 660

gaaacattag atctttgtga tcaaatttta gaaccccttt tagaaaaatc tctcctagat 720

ttaatctttc aagagcaaaa tagccagtta ttagaggaaa cccaaatcac tcagccggtg 780

attt 784

<210> SEQ ID NO 145
<211> LENGTH: 784
<212> TYPE: DNA
<213> ORGANISM: Microcystis sp. 205

<400> SEQUENCE: 145

tcccctaatc ggcctttaat tattggttca gtaaaaacga atttaggcca tttagaaggt 60

gctgccggaa ttgctggatt aattaaaaca gttctagctt tacaacacca taaaatacct 120

cctcatcttc actttaaaaa tcctaatccc cgctttgatt ggagttctca tatttttgaa 180

gttcctgtac aaggaaaacc ttggaatatc agcgaacgtg caaggattgc tggagtaagt 240

tcctttggat ttagtggaac gaatgctcat atcattgttg gggaaattga tgctgatttg 300

cctcaagctt cggagaataa tttttatcta ttaccccttt cggctcgttc tgaacaatcc 360

ctgcaagagt tagcgagaag ctatcaagat attttgactg agtctatcaa tttagctgat 420

gtttgtttta ccaccagtac agggcggggg atttttccgc aacgaatctg tattttagca 480

gactcaataa cgacggcaca acgagcatta attgattacc aagatggtga agattctgac 540

tcattaatcc gaccgatttt atcagagact ccgccaaaga tagctttcct cttttctgga 600

caaggctctc aatattctgg catgggagaa actctttata accgagaagt tgtttttaag 660

gaaacattag atctttgtga tcaaatttta gaaccccttt tagaaaaatc tctcctagat 720

ttaatctttc aagagcaaaa tagccagtta ttagaggaaa cccaaatcac tcagccggtg 780

attt 784

<210> SEQ ID NO 146
<211> LENGTH: 784

-continued

<212> TYPE: DNA
<213> ORGANISM: Planktothrix sp. NIVA-CYA 127

<400> SEQUENCE: 146

cctccgaatc aacctttagt tattggttct gtcaaaacaa atttaggaca cttagaagga 60

gcagcaggaa ttgccggatt aattaaaaca gttttagctt tacaacatca taaaattcct 120

ccccatcttc actttaaaaa acccaaccct cggttggatg ggagttctaa tatttttgaa 180

gttcctgtag gcggaaaacc ctggaatccc agtgaacgcc aaagaattgc cggggtaagt 240

tcctttggct ttagtggaac aaatgctcat attattgtgg gagaaattga ctctagttta 300

cctaaaaaat ctgagcctaa cttttaccta ttaccgcttt cggctcgttc tgaaaaatct 360

ctccaagagt taactaaaaa ttatcaaaat gctttgaatg ggtctgtcaa ttttgctgat 420

gttggtttta cggctactac aggacgggct atttttcagc atcgaatatg tattttagct 480

gaatcaatga caacagcaca ggcagcactg gttagtttcc aaaaaggtga aaattctcaa 540

catttaatta caccaatttt atcagaaaac aagctaaaaa tagcttttct attttcagga 600

caaggctcac aatattcgga aatgggagaa acccttttatc accgagaacc tgtctttaaa 660

aatactttag atatttgtaa tgaaatccta gaacctattt tagaaaaatc cctgttagat 720

ttaatcttta aattgcccaa tagccagcta ttagaacaga ctcaaatcac ccagcccgtg 780

ctat 784

<210> SEQ ID NO 147
<211> LENGTH: 784
<212> TYPE: DNA
<213> ORGANISM: Planktothrix sp. NIVA-CYA 128/R

<400> SEQUENCE: 147

cctccgaatc aacctttagt tattggttct gtcaaaacaa atttaggaca cttagaagga 60

gcagcaggaa ttgccggatt aattaaaaca gttttagctt tacaacacca taaaattcct 120

ccccatcttc actttaaaaa acccaaccct cggttggatg ggagttctaa tatttttgaa 180

gttcctgtag gcggaaaaac ctggaatccc agtgaacgcc aaagaattgc cggggtaagt 240

tcctttggct ttagtggaac aaatgctcat attattgtgg gagaaattga ctctagttta 300

cctaaaaaat ctgagcctaa cttttaccta ttaccgcttt cggctcgttc tgaaaaatct 360

ctccaagagt taactaaaaa ttatcaaaat gctttgaatg ggtctgtcaa ttttgctgat 420

gttggtttta cggctactac aggacgggct atttttcagc atcgaatatg tattttagct 480

gaatcaatga caacagcaca ggcagcactg gttagtttcc aaaaaggtga ggattctcaa 540

catttaatta caccaatttt atcagaaaac aagctaaaaa tagcttttct attttcagga 600

caaggatcgc aatattcagg aatgggagaa acccttttatc accgagaacc tgtctttaaa 660

aatactttag atctttgtaa tgaaatccta gaacctattt tagaaaaatc cctgttagat 720

ttaatcttta aattgcccaa tagccagcta ttagaacaga ctcaaatcac ccagcccgtg 780

ctat 784

<210> SEQ ID NO 148
<211> LENGTH: 784
<212> TYPE: DNA
<213> ORGANISM: Planktothrix 49

<400> SEQUENCE: 148

-continued

cctccgaatc aacctttagt tattggttct gtcaaaacaa atttaggaca cttagaagga 60

gcagcaggaa ttgccggatt aattaaaaca gttttagctt tacaacatca taaaattcct 120

ccccatcttc actttaaaaa acccaaccct cggtttgatg ggagttctaa tatttttgaa 180

gttcctgtag gcggaaaacc ctggaatccc agtgaacgcc aaagaattgc cggggtaagt 240

tcctttggct ttagtggaac aaatgctcat attattgtgg gagaaattga ctctagttta 300

cctaaaaaat ctgagcctaa cttttaccta ttaccgcttt cggctcgttc tgaaaaatct 360

ctccaagagt taactaaaaa ttatcaaaat gctttgaatg ggtctgtcaa ttttgctgat 420

gtttgtttta cggctactac aggacgggct atttttcagc atcgaatatg tattttagct 480

gaatcaatga caacagccca agcagcactg gttagtttcc aaaaaggtga aaattctcaa 540

caattaatta caccaatttt atcagaaaac aagctaaaaa tagcttttct attttcagga 600

caaggctcac aatattcggg aatgggagaa acccttttatc accgagaacc tgtctttaaa 660

aatactttag atatttgtaa tgaaatccta gaacctattt tagaaaaatc cctgttagat 720

ttaatcttta aattgcccaa tagccagcta ttagaacaga ctcaaatcac ccagcccgtg 780

ctat 784

<210> SEQ ID NO 149
<211> LENGTH: 784
<212> TYPE: DNA
<213> ORGANISM: Planktothrix sp. NIVA-CYA 126/8

<400> SEQUENCE: 149

cctccgaatc aacctttagt tattggttct gtcaaaacaa atttaggaca cttagaagga 60

gcagcaggaa ttgccggatt aattaaaaca gttttagctt tacaacacca taaaattcct 120

ccccatcttc actttaaaaa acccaaccct cggtttgatt ggagttctaa tatttttgaa 180

gttcctgtag gcggaaaaac ctggaatccc agtgaacgcc aaggaattgc cggggtaagt 240

tcctttggct ttagtggaac aaatgctcat attattgtgg gagaaattga ctctagttta 300

cctaaaaaat ctgagcctaa cttttaccta ttaccgcttt cggctcgttc tgaaaaatct 360

ctccaagagt taactaaaaa ttatcaaaat gctttgaatg ggtctggcaa ttttgctgat 420

gtttgtttta cgggtactac aggacgggct atttttcagc atcgaatatg tattttagct 480

gaatcaatga caacagcaca agcagcactg gttagtttcc aaaaaggtga ggattctcaa 540

caattaatta caccaatttt atcagaaaac aagctaaaaa tagcttttct attttcagga 600

caaggatcgc aatattcagg aatgggagaa gcccttttatc accgagaacc tgtctttaaa 660

aatactttag atctttgtaa tgaaatccta gaacctattt tagaaaaatc cctgttagat 720

ttaatcttta aattgcccaa tagccagcta ttagaacaga ctcaaatcac ccagcccgtg 780

ctat 784

<210> SEQ ID NO 150
<211> LENGTH: 24
<212> TYPE: DNA
<213> ORGANISM: artificial sequence
<220> FEATURE:
<223> OTHER INFORMATION: discriminating probe; chemically synthetized

<400> SEQUENCE: 150

gcgtaaagag tccgtaggtg gaag 24

-continued

<210> SEQ ID NO 151
<211> LENGTH: 26
<212> TYPE: DNA
<213> ORGANISM: artificial sequence
<220> FEATURE:
<223> OTHER INFORMATION: discriminating probe; chemically synthetized

<400> SEQUENCE: 151

gatgtgatgc gaatctcata aaccgg 26

<210> SEQ ID NO 152
<211> LENGTH: 24
<212> TYPE: DNA
<213> ORGANISM: artificial sequence
<220> FEATURE:
<223> OTHER INFORMATION: discriminating probe; chemically synthetized

<400> SEQUENCE: 152

gcgaaagcgt tctgctagac ctgt 24

<210> SEQ ID NO 153
<211> LENGTH: 27
<212> TYPE: DNA
<213> ORGANISM: artificial sequence
<220> FEATURE:
<223> OTHER INFORMATION: discriminating probe; chemically synthetized

<400> SEQUENCE: 153

ccaaggcttg acatgtcacg aatccta 27

<210> SEQ ID NO 154
<211> LENGTH: 28
<212> TYPE: DNA
<213> ORGANISM: artificial sequence
<220> FEATURE:
<223> OTHER INFORMATION: discriminating probe; chemically synthetized

<400> SEQUENCE: 154

gagcagtgga aactacaaag ctagagtt 28

<210> SEQ ID NO 155
<211> LENGTH: 23
<212> TYPE: DNA
<213> ORGANISM: artificial sequence
<220> FEATURE:
<223> OTHER INFORMATION: discriminating probe; chemically synthetized

<400> SEQUENCE: 155

tcacgctcaa cgtgatcaag gcg 23

<210> SEQ ID NO 156
<211> LENGTH: 28
<212> TYPE: DNA
<213> ORGANISM: artificial sequence
<220> FEATURE:
<223> OTHER INFORMATION: discriminating probe; chemically synthetized

<400> SEQUENCE: 156

catgtcacga attccgttga aagatgga 28

<210> SEQ ID NO 157
<211> LENGTH: 27

-continued

<212> TYPE: DNA
<213> ORGANISM: artificial sequence
<220> FEATURE:
<223> OTHER INFORMATION: common probe; chemically synthetized

<400> SEQUENCE: 157

ttcaagtctg cggtcaaaga atggagg 27

<210> SEQ ID NO 158
<211> LENGTH: 25
<212> TYPE: DNA
<213> ORGANISM: artificial sequence
<220> FEATURE:
<223> OTHER INFORMATION: common probe; chemically synthetized

<400> SEQUENCE: 158

agctcagttc agatcgaagg ctgcg 25

<210> SEQ ID NO 159
<211> LENGTH: 26
<212> TYPE: DNA
<213> ORGANISM: artificial sequence
<220> FEATURE:
<223> OTHER INFORMATION: common probe; chemically synthetized

<400> SEQUENCE: 159

actgacactg agggacgaaa gctagg 26

<210> SEQ ID NO 160
<211> LENGTH: 26
<212> TYPE: DNA
<213> ORGANISM: artificial sequence
<220> FEATURE:
<223> OTHER INFORMATION: common probe; chemically synthetized

<400> SEQUENCE: 160

ttgaaagatg ggagtgcctt cgggat 26

<210> SEQ ID NO 161
<211> LENGTH: 24
<212> TYPE: DNA
<213> ORGANISM: artificial sequence
<220> FEATURE:
<223> OTHER INFORMATION: common probe; chemically synthetized

<400> SEQUENCE: 161

tggtcggggc agaaggaatt cctc 24

<210> SEQ ID NO 162
<211> LENGTH: 25
<212> TYPE: DNA
<213> ORGANISM: artificial sequence
<220> FEATURE:
<223> OTHER INFORMATION: common probe; chemically synthetized

<400> SEQUENCE: 162

gtggaaacta caaagctaga gtgtg 25

<210> SEQ ID NO 163
<211> LENGTH: 23
<212> TYPE: DNA
<213> ORGANISM: artificial sequence
<220> FEATURE:
<223> OTHER INFORMATION: common probe; chemically synthetized

-continued

<400> SEQUENCE: 163

agtgccttcg ggagcgtgaa caa 23

<210> SEQ ID NO 164
<211> LENGTH: 25
<212> TYPE: DNA
<213> ORGANISM: artificial sequence
<220> FEATURE:
<223> OTHER INFORMATION: C-zip code; chemically synthetized

<400> SEQUENCE: 164

gatcggccgg tgaagcgaaa ggttc 25

<210> SEQ ID NO 165
<211> LENGTH: 25
<212> TYPE: DNA
<213> ORGANISM: artificial sequence
<220> FEATURE:
<223> OTHER INFORMATION: C-zip code; chemically synthetized

<400> SEQUENCE: 165

gatggtgatc ccgcgcgtgc cgaaa 25

<210> SEQ ID NO 166
<211> LENGTH: 25
<212> TYPE: DNA
<213> ORGANISM: artificial sequence
<220> FEATURE:
<223> OTHER INFORMATION: C-zip code; chemically synthetized

<400> SEQUENCE: 166

ggattgcacc gtcagcacca ccgag 25

<210> SEQ ID NO 167
<211> LENGTH: 25
<212> TYPE: DNA
<213> ORGANISM: artificial sequence
<220> FEATURE:
<223> OTHER INFORMATION: C-zip code; chemically synthetized

<400> SEQUENCE: 167

tcccaggacg gcgctggcac gttga 25

<210> SEQ ID NO 168
<211> LENGTH: 25
<212> TYPE: DNA
<213> ORGANISM: artificial sequence
<220> FEATURE:
<223> OTHER INFORMATION: C-zip code; chemically synthetized

<400> SEQUENCE: 168

cggcgtccac gtcgagttcc ttcgc 25

<210> SEQ ID NO 169
<211> LENGTH: 25
<212> TYPE: DNA
<213> ORGANISM: artificial sequence
<220> FEATURE:
<223> OTHER INFORMATION: C-zip code; chemically synthetized

<400> SEQUENCE: 169

tgtgcgcccg agatcggtat ccccg 25

-continued

<210> SEQ ID NO 170
<211> LENGTH: 25
<212> TYPE: DNA
<213> ORGANISM: artificial sequence
<220> FEATURE:
<223> OTHER INFORMATION: C-zip code; chemically synthetized

<400> SEQUENCE: 170

atcgcatcgt gatggcgtaa gctcc 25

1. A method for detecting toxic cyanobacteria, characterized in that the method comprises that nucleic acid from a biological sample is brought into contact with an oligonucleotide designed to be specific for the mcyE gene, and the presence or absence of toxic cyanobacteria is detected by a suitable molecular biology method.

2. The method according to claim 1, wherein the oligonucleotide is designed to be specific for a region of the mcyE gene responsible for adding Adda and D-glutamate to the immature synthesis product of microcystin.

3. The method according to claim 1, wherein the oligonucleotide is designed to be specific for a region of the mcyE gene comprising two domains, the adenylation domain and the domain which catalyses a peptide bond between Adda-D-glutamate dipeptide and dehydroalanine.

4. The method according to claim 1, wherein the oligonucleotide is designed to be specific for a fragment of the mcyE gene selected from the group of genera *Anabaena, Microcystis, Planktothrix, Nostoc* and *Nodularia.*

5. The method according to claim 1, wherein the nucleic acid from a biological sample is DNA or RNA.

6. The method according to claim 1, wherein the oligonucleotide is designed to be specific for a fragment of the mcyE gene selected from the group of sequences SEQ ID NO. 1 to SEQ ID NO: 34 as shown in FIG. **19** A to H or to a fragment of said sequences.

7. The method according to claim 1, wherein the oligonucleotide is designed to be specific for a fragment of the mcyE gene selected from the group of consensus sequences SEQ ID NO: 35 to SEQ ID NO: 39 as shown in FIG. **15** A to C or to a fragment of said sequences.

8. The method according to claim 1, wherein the oligonucleotide is selected from the group of mcyE-F2 (SEQ ID NO: 64), AnamcyE-12R (SEQ ID NO: 65) and MicmcyE-R8 (SEQ ID NO:66).

9. The method according to claim 1, wherein the oligonucleotide is selected from the group of discriminating probes SEQ ID NO: 40 to SEQ ID NO: 45.

10. The method according to claim 1, wherein the oligonucleotide is selected from the group of common probes SEQ ID NO: 46 to SEQ ID NO: 51.

11. The method according to claim 1, wherein the detection is combined with a detection method using oligonucleotides designed to be specific for any other mcy gene, such as mcyA or mcyD, or for 16S rRNA.

12. The method according to claim 1, wherein detection is combined with a detection method selected from the group of measuring microcystin concentration, determining cell number, cell density or determining biomass.

13. A fragment of the mcyE gene, characterized in that it is on the region of the mcyE gene responsible for adding Adda and D-glutamate to the immature synthesis product of microcystin and that it is or is located in any of the sequences selected from the group comprising SEQ ID NO. 1 to SEQ ID NO: 34 as shown in FIG. **19** A to H, or is a sequence having at least 80% identity, preferably 90% identity to the sequence.

14. A fragment of the mcyE gene, characterized in that it is on the region of the mcyE gene responsible for adding Adda and D-glutamate to the immature synthesis product of microcystin and that it is or is located in any of the sequences selected from the group comprising consensus sequences SEQ ID NO: 35 to SEQ ID NO: 39 as shown in FIG. **15** A to C, or is a sequence having at least 80% identity, preferably 90% identity to the sequence.

15. An oligonucleotide, characterized in that it is designed to be specific for the region of the mcyE gene responsible for adding Adda and D-glutamate to the immature synthesis product of microcystin that is or is located in any of the sequences selected from the group comprising SEQ ID NO: 1 to SEQ ID NO: 34 as shown in FIG. **19** A to H or selected from the group comprising any of the consensus sequences SEQ ID NO: 35 to SEQ ID NO: 39 as shown in FIG. **15** A to C.

16. An oligonucleotide selected from the group of mcyE-F2 (SEQ ID NO: 64), AnamcyE-12R (SEQ ID NO: 65) and MicmcyE-R8 (SEQ ID NO:66).

17. An oligonucleotide selected from the group of discriminating probes of SEQ ID NO: to SEQ ID NO: 45.

18. An oligonucleotide selected from the group of common probes of SEQ ID NO: 46 to SEQ ID NO: 51.

19. mcyE gene from *Anabaena* genus encoding the amino acid sequence of SEQ ID NO: or a sequence having at least 80% identity, preferably 90% identity to the sequence, or a fragment of said sequence having polymorphic sites which make possible of designing oligonucleotides to be specific for the fragment.

20. mcyE gene from *Anabaena* genus having the nucleic acid sequence SEQ ID NO: 68 or a sequence having at least 80% identity, preferably 90% identity to the sequence, or a fragment of said sequence having polymorphic sites, which make possible of designing oligonucleotides to be specific for the fragment.

21. The method according to claim 12, wherein the detection is combined with a detection method using oligonucleotides designed to be specific for a fragment of the mcyD gene which is on the region of the mcyD gene responsible for chain elongation of the growing Adda amino

acid in the synthesis of microcystin and that it is or is located in any of the sequences selected from the group comprising sequences SEQ ID NO: 131 to SEQ ID NO: 149 as shown in FIG. **38** A to F or is a sequence having at least 85% identity, preferably 90% identity to the sequence.

22. The method according to claim 12, wherein the detection is combined with a detection method using oligonucleotides designed to be specific for mcyD gene from *Anabaena* genus encoding the amino acid sequence of SEQ ID NO: 69 or a sequence having at least 80% identity, preferably 90% identity to the sequence, or a fragment of said sequence having polymorphic sites which make possible of designing oligonucleotides to be specific for the fragment.

23. The method according to claim 12, wherein the detection is combined with a detection method using oligonucleotides designed to be specific for mcyD gene from *Anabaena* genus having the nucleic acid sequence SEQ ID NO: 70 or a sequence having at least 80% identity, preferably 90% identity to the sequence or a fragment of said sequence having polymorphic sites which make possible of designing oligonucleotides to be specific for the fragment.

24. An oligonucleotide selected from the group of discriminating probes of SEQ ID NO:71 to SEQ ID NO:90.

25. An oligonucleotide selected from the group of common probes of SEQ ID NO:91 to SEQ ID NO:110.

26. An oligonucleotide selected from the group of discriminating probes of SEQ ID NO:150 to SEQ ID NO:163.

27. An oligonucleotide selected from the group of common probes of SEQ ID NO:157 to SEQ ID NO:163.

28. A kit for detection of toxic cyanobacteria by microarray method, characterized in that it comprises

discriminating probes and common probes designed to be specific for mcy-E gene, and optionally for mcyD gene;

DNA or RNA zip and complementary zip codes assigned to be specific for selected cyanobacteria genera.

29. A kit for detection of toxic cyanobacteria by hybridization, characterized in that it comprises

primers designed to be specific for the mcyE gene, and optionally for mcyD gene;

probes designed to be specific for selected cyanobacteria genera.

30. The kit according to claim 29, which comprises in addition to primers and probes designed to be specific for mcy gene also primers and probes for 16S rRNA gene.

31. A method for detecting toxic and non-toxic cyanobacteria, characterized in that the method comprises that nucleic acid from a biological sample is brought into contact with an oligonucleotide designed to be specific for mcyE gene or for other mcy genes, such as mcyD gene, and with an oligonucleotide designed to be specific for the 16SrRNA gene, and the presence or absence of toxic cyanobacteria is detected by a suitable molecular biology method.

32. The method according to claim 31, wherein the oligonucleotides are designed to be specific for a region of the mcyE and for a region of 16SrRNA gene.

33. The method according to claim 1, wherein the molecular biology method is selected from the group comprising hybridization, PCR, reverse transcriptase PCR, QTR-PCR, LCR, LDR and minisequencing.

34. The method according to claim 1, wherein the detection is made by microarray method.

35. The kit according to claim 28, which comprises in addition to primers and probes designed to be specific for mcy gene also primers and probes for 16SrRNA gene.

36. The method according to claim 31, wherein the molecular biology method is selected form the group comprising hybridization, PCR, reverse transcriptase PCR, QTR-PCR, LCR, LDR and minisequencing.

37. The method according to claim 31, wherein the detection is made by microarray method.

* * * * *