Scientific Reports—Article

Title:

Genomic divergence within non-photosynthetic cyanobacterial endosymbionts in rhopalodiacean diatoms

Authors: Takuro Nakayama* and Yuji Inagaki

Center for Computational Sciences, University of Tsukuba, 1-1-1 Tennoudai, Tsukuba, Ibaraki 305-8577, Japan

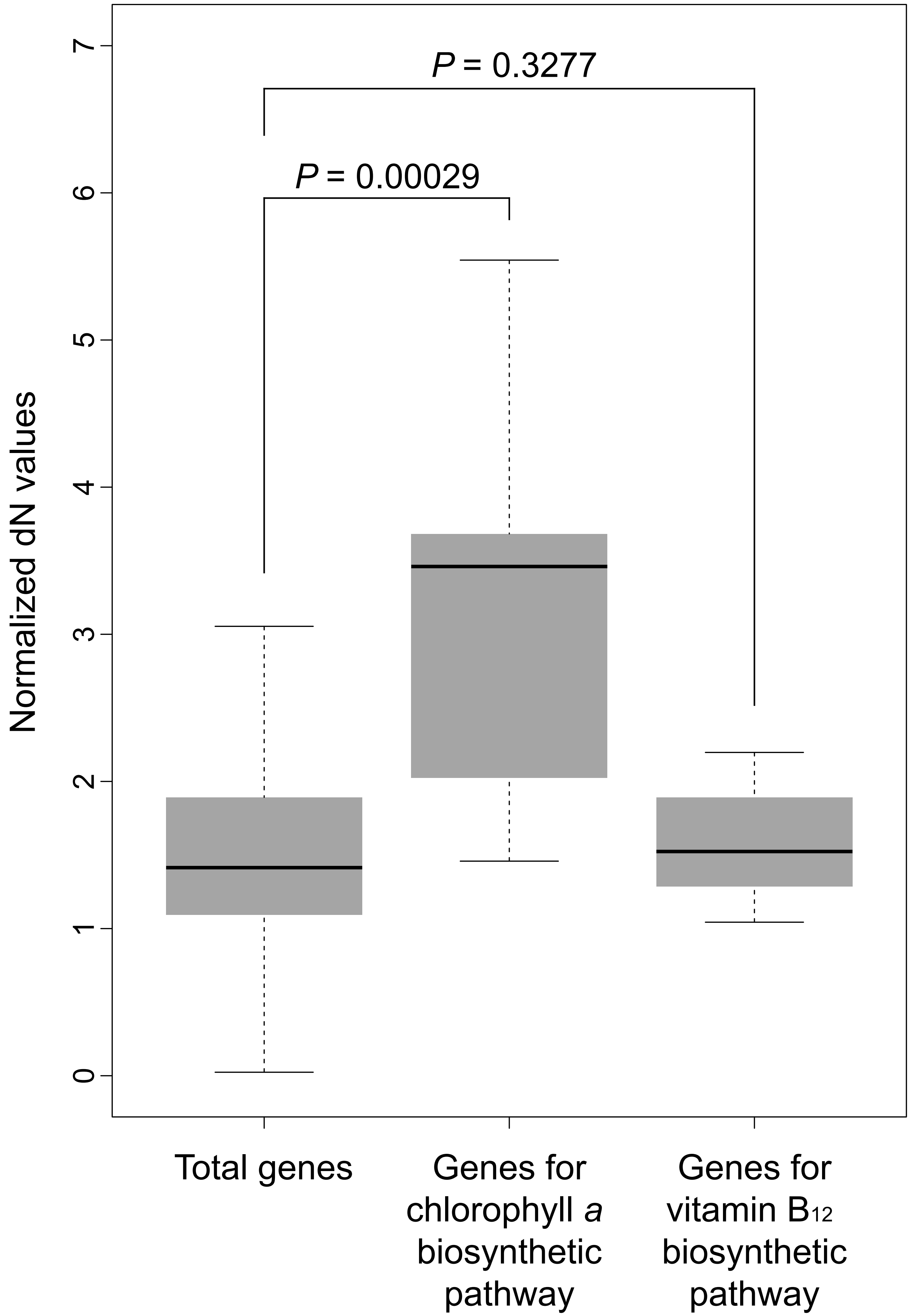
Supplementary Figure Legends

Figure S1. Normalized rates of non-synonymous substitution for total gene set and genes for the chlorophyll *a* and vitamin B_{12} biosynthetic pathways. Each boxplot represents the distribution of non-synonymous substitution rates (dN values) on protein-coding genes in the *Rg*SB genome, which were normalized by that of counterpart genes from *Cyanothece* spp. genomes (see material and method section in the main text). The difference in the normalized dN values between the total gene set and the Chl-*a* biosynthetic genes appeared to be significant (the Wilcoxon rank-sum test, $P = 0.00029$). On the other hand, no significant difference in the dN value distribution was observed in the same test considering the genes for vitamin B_{12} biosynthesis (the Wilcoxon rank-sum test, $P = 0.3277$). Outliers were not displayed in this graph.

Figure S2. Sequence comparison of the methionine synthase from the *Rg*SB, *Et*SB and free-living cyanobacteria. (a) Multiple-alignment of the amino acid sequences. (b) Pairwise identity matrix of sequences displayed in (a). The values indicate percentages of amino acid identity.

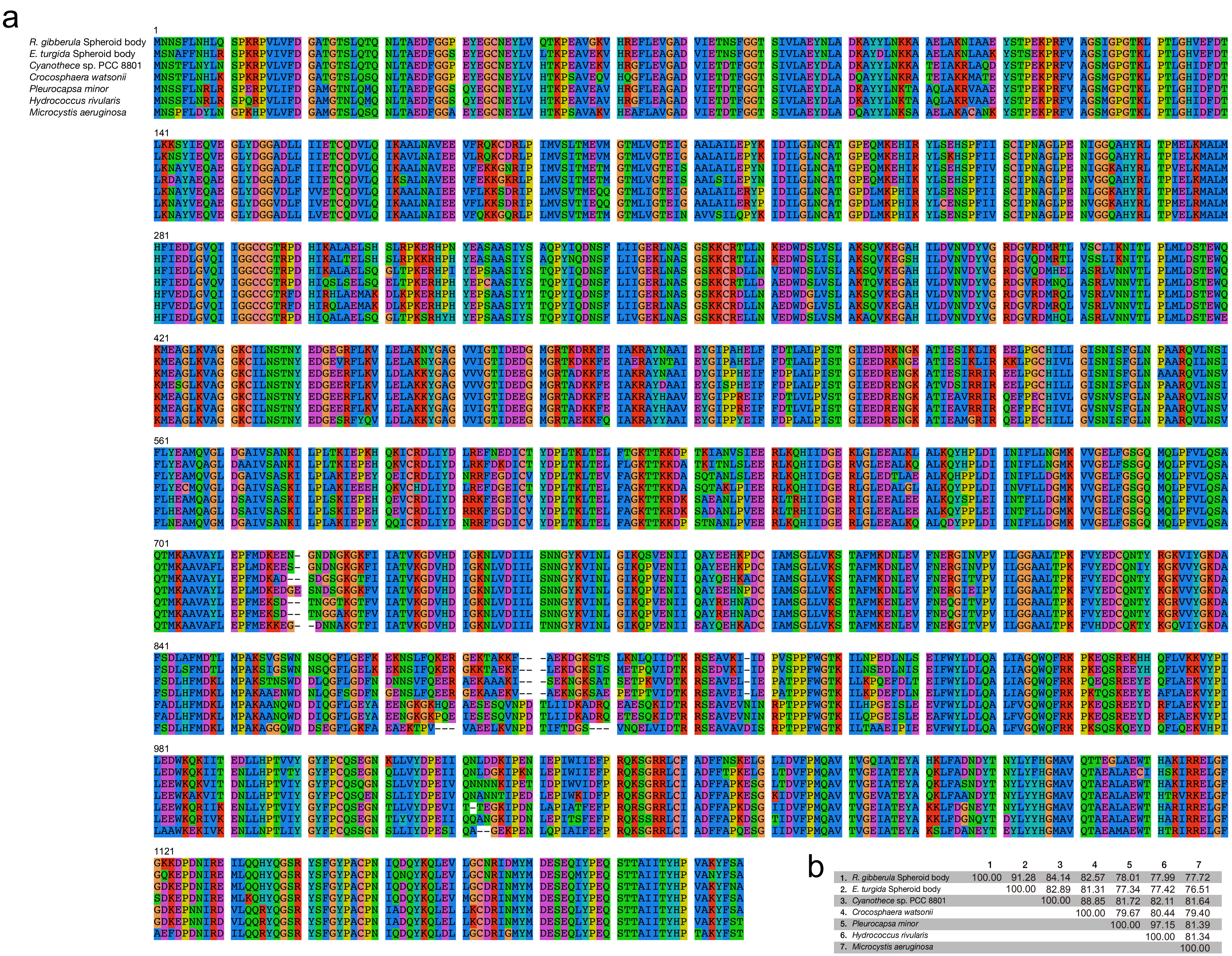
Figure S3. Phylogenetic tree topologies used in CODEML analyses.

Figure S1



P = 0.3277
P = 0.00029
Normalized dN values
0
1
2
3
4
5
6
7
Total genes
Genes for chlorophyll a biosynthetic pathway
Genes for vitamin B12 biosynthetic pathway

Figure S2



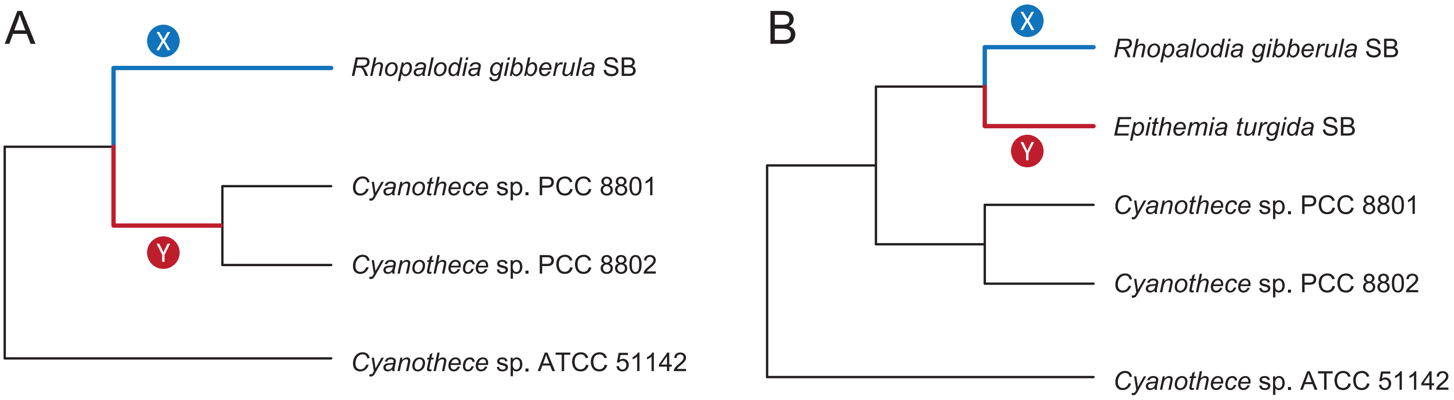
a

```
                              1
R. gibberula Spheroid body    MNNSFLNHLQ SPKRPVLVFD GATGTSLQTQ NLTAEDFGGP EYEGCNEYLV QTKPEAVGKV HREFLEVGAD VIETNSFGGT SIVLAEYNLA DKAYYLNKKA AELAKNIAAE YSTPEKPRFV AGSIGPGTKL PTLGHVEFDT
E. turgida Spheroid body      MSNAFFNHLR SPKRPVIVFD GATGTSLQTQ NLTAEDFGGS EYEGCNEYLV LTKPEAVEKV HREFLEAGAD VIETNSFGGT SIVLAEYNLA DKAYDLNKKA AELAKNLAAK YSTSEKPRFV AGSIGPGTKL PTLGHVEFDT
Cyanothece sp. PCC 8801       MNSTFLNYLN SPKRPVLVFD GATGTSLQTQ NLTAEDFGGP EYEGCNEYLV HTKPEAVEKV HRGFLEVGAD VIETDTFGGT SIVLAEYDLA DKAYYLNKKA TEIAKRLAQD YSTPEKPRFV AGSMGPGTKL PTLGHIDFDT
Crocosphaera watsonii         MNSTFLNHLK SPKRPVLVFD GATGTSLQSQ NLTAEDFGGP EYEGCNEYLV HTKPSAVEQV HQGFLEAGAD VIETDTFGGT SIVLAEYDLA DQAYYLNKRA TEIAKKMATE YSTPEKPRFV AGSMGPGTKL PTLGHIDFDT
Pleurocapsa minor             MNSSFLNRLR SPERPVLIFD GAMGTNLQMQ NLTAEDFGGS QYEGCNEYLV HTKPEAVEAV HRGFLEAGAD VIETDTFGGT SIVLAEYDLA DKAYYLNKTA AQLAKRVAAE YSTPEKPRFV AGSMGPGTKL PTLGHIDFDT
Hydrococcus rivularis         MNSSFLNRLR SPQRPVLIFD GAMGTNLQMQ NLTAEDFGGS QYEGCNEYLV HTKPEAVEAV HRGFLEAGAD VIETDTFGGT SIVLAEYDLA DKAYYLNKTA AQLAKRVAAE YSTPEKPRFV AGSMGPGTKL PTLGHIDFDT
Microcystis aeruginosa        MNSPFLDYLN GPKHPVLVFD GAMGTSLQSQ NLTAEDFGGA EYEGCNEYLV HTKPSAVAKV HEAFLAVGAD VIETDTFGGT SIVLAEYDLA DQAYYLNKSA AELAKACANK YSTPEKPRFV AGSMGPGTKL PTLGHIDFDT

141
LKKSYIEQVE GLYDGGADLL IIETCQDVLQ IKAALNAVEE VFRQKCDRLP IMVSLTMEVM GTMLVGTEIG AALAILEPYK IDILGLNCAT GPEQMKEHIR YLSEHSPFII SCIPNAGLPE NIGGQAHYRL TPMELKMALM
LKNSYIEQVE GLYDGGADLL IIETCQDVLQ IKAVLNAIEE VLRQKCDRLP IMVSITMEVM GTMLVGTEIG AALAILEPYK IDILGLNCAT GPEQMKEHIR YLSKHSPFII SCIPNAGLPE NIGGQAHYRL TPIELKMALM
LKNAYVEQAE GLYDGGADLF IIETCQDVLQ IKAALNAVEE VFEKKGNRIP IMVSITMETM GTMLVGTEIS AALAILEPYN IDILGLNCAT GPEQMKEHIK YLSEHSPFIV SCIPNAGLPE NIGGKAHYRL TPVELKMALM
LRDAYAEQAE GLYDGGADLF IIETCQDVLQ IKSALNAVEE VFEKKGKRLP LMVSITMETM GTMLVGTEIS AALSILEPYN IDILGLNCAT GPEQMKEHIK YLSEHSPFII SCIPNAGLPE NVGGQAHYRL TPIELKMALM
LKNAYVEQAE GLYDGGVDLF VVETCQDVLQ IKAALNAIEE VFLKKSDRIP LMVSVTMEQQ GTMLIGTEIG AALAILERYP IDILGLNCAT GPDLMKPHIR YLSENSPFII SCIPNAGLPE NLGGKAHYRL TPMELRMALM
LKNAYVEQAE GLYDGGVDLF IVETCQDVLQ IKAALNAIEE VFLKKSDRIP LMVSVTIEQQ GTMLIGTEIG AALAILERYP IDILGLNCAT GPDLMKPHIR YLCENSPFII SCIPNAGLPE NVGGKAHYRL TPMELRMALM
LKNAYVEQVE GLYDGGADLL LVETCQDVLQ IKAALNAIEE VFQKKGQRLP LMVSVTMETM GTMLVGTEIN AVVSILQPYK IDILGLNCAT GPDLMKPHIK YLSENSPFIV SCIPNAGLPE NVGGQAHYRL TPVELKMALM

281
HFIEDLGVQI IGGCCGTRPD HIKALAELSH SLRPKERHPN YEASAASIYS AQPYIQDNSF LIIGERLNAS GSKKCRTLLN KEDWDSLVSL AKSQVKEGAH ILDVNVDYVG RDGVRDMRTL VSCLIKNITL PLMLDSTEWQ
HFIEDLGVQI IGGCCGTRPD HIKALTELSH SLRPKKRHPH YEASAASIYS AQPYNQDNSF LIIGEKLNAS GSKKCRTLLN KEDWDSLISL AKSQVKEGSH ILDVNVDYVG RDGVQDMRTL VSSLIKNITL PLMLDSTEWQ
HFIEDLGVQI IGGCCGTRPD HIKALAELSQ GLTPKERHPI YEPSAASIYS TQPYIQDNSF LIVGERLNAS GSKKCRDLLN VEDWDSLVSL AKSQVKEGAH ILDVNVDYVG RDGVQDMHEL ASRLVNNVTL PLMLDSTEWQ
HFIEDLGVQV IGGCCGTRPD HIQSLSELSQ ELTPKERHPH YEPCAASIYS TQPYIQDNSF LIIGERLNAS GSKKCRTLLD AEDWDSLVSL AKTQVKEGAH VLDVNVDYVG RDGVRDMNQL ASRLVNNVTL PLMLDSTEWQ
HFVEDLGVQI IGGCCGTRFD HIRHLAEMAK DLKPKERHPH YEPSAASIYT TQPYIQDNSF LIIGERLNAS GSKKCRDLLN AEDWDGLVSL AKSQVKEGAH VLDVNVDYVG RDGVRDMRQL VSRLVNNITL PLMLDSTEWQ
HFVEDLGVQI IGGCCGTRFD HIRQLAEMAK DLKPKERHPH YEPSAASIYT TQPYIQDNSF LIIGERLNAS GSKKCRDLLN AEDWDGLVSL AKSQVKEGAH VLDVNVDYVG RDGVRDMRQL VSRLVNNVTL PLMLDSTEWQ
HFIEDLGVQI IGGCCGTRPD HIQALAELSQ GLTPKSRHYH YEPSAASIYS TQPYIQDNSF LIVGEKLNAS GSKKCRELLN VEDWDSLVSM AKAQVKEGAH ILDVNVDYVG RDGVRDMHQL ASRLVNNVTL PLMLDSTEWE

421
KMEAGLKVAG GKCILNSTNY EDGEGRFLKV LELAKNYGAG VVIGTIDEDG MGRTKDRKFE IAKRAYNAAI EYGIPAHELF FDTLALPIST GIEEDRKNGK ATIESIKLIR EELPGCHILL GISNISFGLN PAARQVLNSI
KMEAGLKVAG GKCILNSTNY EDGEVRFLKV LELAKNYGAG VVIGTIDEDG MGRTKDKKFE IAERAYNTAI EYGIPAHELF FDTLALPIST GIEEDRKNGE ATIESIKLIR KKLPGCHIVL GISNISFGLN PAARQILNSI
KMEAGLKVAG GKCILNSTNY EDGEERFLKV LDLAKKYGAG VVVGTIDEEG MGRTADKKFE IAKRAYNAAI DYGIPPHEIF FDPLALPIST GIEEDRENGK ATIESIRRIR EELPGCHILL GISNISFGLN AAARQVLNSV
KMESGLKVAG GKCILNSTNY EDGEERFLKV LELAKKYGAG VVVGTIDEDG MGRTADKKFE IAKRAYDAAI EYGISPHEIF FDPLALPIST GIEEDRENGK ATVDSIRRIR EELPGCHILL GISNISFGLN PAARQVLNSI
KMEAGLKVAG GKCILNSTNY EDGEERFLKV LELAKKYGAG VVIGTIDEEG MGRTADKKFE IAKRAYHAAV EYGIPPREIF FDTLVLPVST GIEEDRENGK ATIEAVRRIR QEFPECHIVL GVSNVSFGLN PAARQVLNSV
KMEAGLKVAG GKCILNSTNY EDGEERFLKV LELAKKYGAG IVIGTIDEEG MGRTADKKFE IAKRAYHAAV EYGIPPREIF FDTLVLPIST GIEEDRENGK ATIEAVRRIR QEFPECHIVL GVSNVSFGLN PAARQVLNSV
KMEAGLKVAG GKCILNSTNY EDGESRFYQV LDLAKKYGAG VVIGTIDEEG MGRTAEKKFQ IAKRAYYAAI EYGIPPYEIF FDPLALPIST GIEEDRENGK ATIEAMGRIR QELPECHILL GVSNISFGLN PAARQVLNSV

561
FLYEAMQVGL DGAIVSANKI LPLTKIEPKH QKICRDLIYD LREFNEDICT YDPLTKLTEL FTGKTTKKDP TKIANVSIEE RLKQHIIDGE KLGLEEALKL ALKQYHPLDI INIFLLNGMK VVGELFGSGQ MQLPFVLQSA
FLYEAVQAGL DAAIVSANKI LPLTKIEPEH QKVCRDLIYD LRKFDKDICT YDPLTKLTEL FLGKTTKKDA TKITNLSLEE RLKQHIIDGE RVGLEEALKQ ALKQYHPLDI INIFLLNGMK VVGELFSSGQ MQLPFVLQSA
FLYEAMQVGL DGAIVSASKI LPLAKIEPEY QEICRDLIYD NRRFEGDICV YDPLTKLTEV FAGKTTKKDA SQTANLSLEE RLKQHIIDGE RLGLEDTLAE ALKQHPPLDI INIFLLDGMK VVGELFGSGQ MQLPFVLQSA
FLYECMQVGL DGAIVSASKI LPLAKIEEEH QKVCHDLIYD LREFDGEICT YDPLTKLTEL FAGKTTKKDA SQTAKLPIEE RLKQHIIDGE RLGLEDALGL ALKQYPPLDI INVFLLDGMK VVGELFGSGQ MQLPFVLQSA
FLHEAMQAGL DSAIVSASKI LPLSKIEPEH QEVCRDLIYD RRKFEGEICV YDPLTKLTEL FAGKTTKRDK SAEANLPVEE RLTRHIIDGE RIGLEEALEE ALKQYSPLEI INTFLLDGMK VVGELFGSGQ MQLPFVLQSA
FLHEAMQAGL DSAIVSASKI LPLSKIEPEH QEVCRDLIYD RRKFEGDICV YDPLTKLTEL FAGKTTKRDK SADANLPVEE RLTRHIIDGE RIGLEEALEE ALKQYPPLEI INTFLLDGMK VVGELFGSGQ MQLPFVLQSA
FLNEAMQVGM DGAIVSANKI LPLAKIEPEY QQICRDLIYD NRRFDGDICV YDPLTKLTEL FAGKTTKKDP STNANLPVEE RLKQHIIDGE RLGLEEALKQ ALQDYPPLDI INIFLLDGMK VVGELFGSGQ MQLPFVLQSA

701
QTMKAAVAYL EPFMDKEEN- GNDNGKGKFI IATVKGDVHD IGKNLVDIIL SNNGYKVINL GIKQSVENII QAYEEHKPDC IAMSGLLVKS TAFMKDNLEV FNERGINVPV ILGGAALTPK FVYEDCQNTY RGKVIYGKDA
QTMKAAVAFL EPLMDKEES- GNDNGKGKFI IATVKGDVHD IGKNLVDIIL SNNGYKVINL GIKQPVENII QAYEEHKPDC IAMSGLLVKS TAFMKDNLEV FNERGINVPV ILGGAALTPK FVYEDCQNIY KGKVVYGKDA
QTMKAAVAYL EPFMDKAD-- SDGSGKGTFI IATVKGDVHD IGKNLVDIIL SNNGYKVINL GIKQPVENII QAYQEHKADC IAMSGLLVKS TAFMKDNLEV FNERGITVPV ILGGAALTPK FVYEDCQNTY KGKVVYGKDA
QTMKAAVAYL EPLMDKEDGE SNDSGKGKFV IATVKGDVHD IGKNLVDIIL SNNGYKVINL GIKQPVENII QAYEEHNPDC IAMSGLLVKS TAFMKDNLEV FNERGIEIPV ILGGAALTPK FVYEDCQNTY KGKVVYGKDA
QTMKAAVAYL EPFMEKSD-- TNGGTKGTFV IATVKGDVHD IGKNLVDIIL SNNGYKVINL GIKQPVENII QAYREHNADC IAMSGLLVKS TAFMKENLEV FNEQGITVPV ILGGAALTPK FVYEDCQNTY KGRVVYGKDA
QTMKAAVAYL EPFMEKSD-- TNGGAKGTFV IATVKGDVHD IGKNLVDIIL SNNGYKVINL GIKQPVENII QAYREHNADC IAMSGLLVKS TAFMKENLEV FNEQGITVPV ILGGAALTPK FVYEDCQNTY KGRVVYGKDA
QTMKAAVAFL EPFMEKKEG- -DNNAKGTFI IATVKGDVHD IGKNLVDIIL TNNGYRVINL GIKQPVENII EAYQEHKADC IAMSGLLVKS TAFMKENLEV FNEKGITVPV ILGGAALTPK FVHDDCQKTY KGQVIYGKDA

841
FSDLAFMDTL MPAKSVGSWN NSQGFLGEFK EKNSLFQKER GEKTAKKF-- -AEKDGKSTS LKNLQIIDTK RSEAVKI-ID PVSPPFWGTK ILNPEDLNLS EIFWYLDLQA LIAGQWQFRK PKEQSREKHH QFLVKKVYPI
FSDLSFMDTL MPAKSIGSWN NSQGFLGELK EKNSIFQKER GKKTAEKF-- -LEKDGKSIS METPQVIDTK RSEDVKI-ID PVSPPFWGTK ILNSEDLNIS EIFWYLDLQA LIAGQWQFRK PKEQSREEYH QFLVKKVYPI
FSDLHFMDKL MPAKSTNSWD DLQGFLGDFN DNNSVFQEER AEKAAAKI-- -SEKNGKSAT SETPKVVDTK RSEAVEL-IE PATPPFWGTK LLKPQEFDLT EIFWYLDLQA LIAGQWQFRK PQEQSREEYE QFLAEKVYPI
FSDLHFMDKL MPAKAAENWD NLQGFSGDFN GENSLFQEER GEKAAEKV-- -AEKNGKSAE PETPTVIDTK RSEAVEI-LE PATPPFWGTK ILKPDEFDLN ELFWYLDLQA LIAGQWQFRK PKTQSKEEYE AFLAEKVYPI
FADLHFMDKL MPAKAANQWD DIQGFLGEYA EENGKGKHQE AESESQVNPD TLIIDKADRQ EAESQKIDTR RSEAVEVNIN RPTPPFWGTK ILHPGEISLE EVFWYLDLQA LFVGQWQFRK PKEQSREEYD RFLAEKVYPI
FADLHFMDKL MPAKAANQWD DIQGFLGEYA EENGKGKPQE IESESQVNPD TLIIDKADRQ ETESQKIDTR RSEAVEVNIN RPTPPFWGTK ILQPGEISLE EAFWYLDLQA LFVGQWQFRK PKEQSREEYD RFLAEKVYPI
FSDLHFMDKL MPAKAGGQWD DSEGFLGKFA EAEKTPV--- VAEELKVNPD TIFTDGS--- VNQELVIDTR RSEAVAVDIS RPTPPFWGTK ILTAAEIPIE EVFWYLDLQA LFVGQWQFRK PKSQSKQEYD QFLQEKVHPI

981
LEDWKQKIIT EDLLHPTVVY GYFPCQSEGN KLLVYDPEII QNLDDKIPEN IEPIWIIEFP RQKSGRRLCF ADFFNSKELG LIDVFPMQAV TVGQIATEYA HKLFADNDYT NYLYFHGMAV QTTEGLAEWT HAKIRRELGF
LEDWKQKIIT ENLLHPTVTY GYFPCQSEGN QLLVYDPEII QNLDGKIPKN LEPIWIIEFP RQKSGRRLCI ADFFTPKELG LTDVFPMQAV TVGEIATEYA QKLFADNDYT NYLYFHGMAV QTAEALAECI HSKIRRELGF
LEEWKQKVIT DNLLHPTVIY GYFPCQSQGN SLLVYDPEIV QNNNNKIPET IDPIWVIDFP RQKSGRRLCI ADFFAPKESG LTDVFPMQAV TVGEIATEYA QKLFASNDYT NYLYYHGMAV QTAEALAEWT HAKIRRELGF
LEEWKAKVIT DNLLHPTVIY GYFPCQSQEN SLLIYDPEIV QNANNTIPED LEPIWKIDFP RQKSGRRLCI ADFFAPKESG KIDVFPMQAV TVGEVATEYA QKLFAANDYT NYLYYHGMAV QTAEALAEWT HTRVRKELGF
LEEWKQRIIK ENLLYPTVIY GYFPCQSEGN TLLVYDPEVI T-TEGKIPDN LAPIATFEFP RQRSGRRLCI ADFFAPKDSG IIDVFPMQAV TVGEIATEYA KKLFDGNEYT NYLYYHGMAV QTAEALAEWT HARIRRELGF
LEEWKQRIVK ENLLHPTVIY GYFPCQSEGN TLYVYDPEII QQANGKIPDN LEPITSFEFP RQKSSRRLCI ADFFAPKDSG TIDVFPMQAV TVGEIATEYA KKLFDGNQYT DYLYYHGMAV QTAEALAEWT HARIRRELGF
LAAWKEKIVK ENLLNPTLIY GYFPCQSSGN SLLIYDPESI QA--GEKPEN LQPIAIFEFP RQKSGRRLCI ADFFAPQESG IIDVFPMQAV TVGEIATEYA KSLFDANEYT EYLYYHGMAV QTAEAMAEWT HTRIRRELGF

1121
GKKDPDNIRE ILQQHYQGSR YSFGYPACPN IQDQYKQLEV LGCNRIDMYM DESEQIYPEQ STTAIITYHP VAKYFSA
GQKEPDNIRE MLQQHYQGSR YSFGYPACPN IQNQYKQLEV LGCNRINMYM DESEQMYPEQ STTAIITYHP VANYFSA
GDKEPDNIRE MLQQHYQGSR YSFGYPACPN IQDQYKQLEI LGCDRINMYM DESEQIYPEQ STTAIITYHP VAKYFNA
SDKEPDNIRE MLKQHYQGSR YSFGYPACPN IQDQYKQLEV MGCDRINMYM DESEQIYPEQ STTAIITYHP VAKYFSA
GDKEPNNIRD VLQQRYQGSR YSFGYPACPN IADQYKQLEL LKCDRINMYM DESEQLYPEQ STTAIITYHP VAKYFSA
GDKEPNNIRD VLQQRYQGSR YSFGYPACPN IADQYKQLEL LKCDRINMYM DESEQLYPEQ STTAIITYHP VAKYFST
AEFDPDNIRD ILQQRYQGSR YSFGYPACPN IQDQYKQLDL LGCDRIGMYM DESEQLYPEQ STTAIITYHP TAKYFST
```

b

	1	2	3	4	5	6	7
1. *R. gibberula* Spheroid body	100.00	91.28	84.14	82.57	78.01	77.99	77.72
2. *E. turgida* Spheroid body		100.00	82.89	81.31	77.34	77.42	76.51
3. *Cyanothece* sp. PCC 8801			100.00	88.85	81.72	82.11	81.64
4. *Crocosphaera watsonii*				100.00	79.67	80.44	79.40
5. *Pleurocapsa minor*					100.00	97.15	81.39
6. *Hydrococcus rivularis*						100.00	81.34
7. *Microcystis aeruginosa*							100.00

Figure S3


A
X
Rhopalodia gibberula SB
Cyanothece sp. PCC 8801
Y
Cyanothece sp. PCC 8802
Cyanothece sp. ATCC 51142
B
X
Rhopalodia gibberula SB
Epithemia turgida SB
Y
Cyanothece sp. PCC 8801
Cyanothece sp. PCC 8802
Cyanothece sp. ATCC 51142