*Scientific paper*

# Discrimination Between *Synechocystis* Members (Cyanobacteria) Based on Heterogeneity of Their 16S rRNA and ITS Regions

**Mojca Juteršek, Marina Klemenčič and Marko Dolinar***

*Department of Chemistry and Biochemistry, Faculty of Chemistry and Chemical Technology, University of Ljubljana, Večna pot 113, SI-1000 Ljubljana, Slovenia*

** Corresponding author: E-mail: marko.dolinar@fkkt.uni-lj.si*

*Received: 06-02-2017*

## Abstract

Cyanobacteria are an important group of microorganisms displaying a range of morphologies that enable phenotypic differentiation between the major lineages of cyanobacteria, often to the genus level, but rarely to species or strain level. We focused on the unicellular genus *Synechocystis* that includes the model cyanobacterial strain PCC 6803. For 11 *Synechocystis* members obtained from cell culture collections, we sequenced the variable part of the 16S rRNA-encoding region and the 16S - 23S internally transcribed spacer (ITS), both standardly used in taxonomy. In combination with microscopic examination we observed that 2 out of 11 strains from cell culture collections were clearly different from typical *Synechocystis* members. For the rest of the samples, we demonstrated that both sequenced genomic regions are useful for discrimination between investigated species and that the ITS region alone allows for a reliable differentiation between *Synechocystis* strains.

**Keywords:** Synechocystis; DNA barcoding; Cyanobacteria; rRNA; ITS region

## 1. Introduction

Cyanobacteria are Gram-negative prokaryotes characterized by their ability to execute oxygenic photosynthesis. They inhabit various environments, from oceans to freshwaters, but also including extreme locations such as deserts, hot springs and hypersaline habitats.[1] As a consequence, there is a considerable morphological diversity among these organisms, which was traditionally the key for taxonomic classification of cyanobacteria. However, improper growth conditions of wild strains when transferred to laboratory environment may result in the loss of morphological characteristics[2,3] which consequently leads to misidentification and false classification.

To overcome variable morphological criteria, DNA-based methods are becoming widely applied in the identification and cataloguing of cyanobacteria, either as the sole method of identification or in combination with phenotypic and ecological characterization.[4] Adherent to classification of other bacteria, DNA-based taxonomy in cyanobacteria is mostly based on similarity in their 16S rRNA sequences, with the assumption that individuals of the same species share greater sequence similarity than individuals of different species.[5] Although overall evolution of the 16S rRNA gene is rather slow, there are regions that are more variable, which allows for studying evolutionary relationships both between distant and closely related groups of organisms.[6,7]

Phylogenetic analysis based on 16S rRNA relies on the presumption that its gene only occurs in one copy per genome, or in case of multiple rRNA genes, that they are identical in sequence. Cyanobacteria commonly contain multiple ribosomal RNA operons and point–mutations can often be found in paralogous 16S rRNA gene copies. But since sequence heterogeneity is relatively low (mean = 0.2%), it is believed to have no significant impact on determining phylogenetic relationships.[8] Although the use of 16S rRNA gene sequences remains a common tool for identification of organisms to the species level, doubts were expressed whether there is sufficient variability in 16S rRNA gene sequences to allow for discrimination at the subgeneric level.[9]

Owing to increasing number of sequenced cyanobacterial genomes, which has already exceeded the number of 150,[10] the current phylogenetic studies that are in part based on 16S rRNA, also include a selection of more

variable sequences. In addition to sequences of protein-coding genes, e.g. *psbA*, *rbcL*, *rnpB*, *rpoC*, *gyrB*,[11,12] research has increasingly focused on the internal transcribed spacer of ribosomal RNA genes (16S–23S rRNA-ITS).[13] With its variable length and number,[14] rRNA ITS region is becoming a popular tool in identification and classification of cyanobacteria.[15] Three types of ITS regions were identified up to now in cyanobacteria, differing in the presence or absence of specific tRNA genes (reviewed by Sarma,[16]): the first type contains both $tRNA^{Ile}$ and $tRNA^{Ala}$ coding sequences (as found in *Anabaena* sp., *Nostoc* sp. or *Synechococcus* sp. PCC 6301), the second type contains only $tRNA^{Ile}$ (found e.g. in 47 strains of *Microcystis*, in *Synechocystis* sp. PCC 6803 and *Spirulina* sp. PCC 6313), while the third type has no identifiable tRNA-encoding sequence (as found in *Nodularia* sp. BCNO D9427). Restriction endonuclease digestion of amplified rRNA-ITS genomic segments has been used to delineate closely related cyanobacterial strains,[17] whereas sequencing has been shown to be successful in analysis of subgeneric relationships of *Microcystis,*[18,19] *Trichodesmium,*[20] *Synechococcus,*[15,21] *Prochlorococcus,*[22] *Aphanizomenon* and *Anabaena*[3] as well as various picocyanobacteria.[23]

Surprisingly, no in-depth taxonomic classification has been performed for the genus *Synechocystis.*[24] Although more than 20 species have been described and many more strains were deposited in culture collections, limited sequence data as well as lack of details at the subcellular level hinder adequate identification and classification. Several planktic species including *S. salina*, *S. limnetica*, *S. aquatilis*, and a few picoplanktic types are hardly morphologically distinguishable,[25] which calls for a molecular biological approach.

In our study, 11 different *Synechocystis* representatives were analysed for their 16S rRNA and ITS sequence properties. Up to now, 16S rRNA data were available only for a few strains, most of them not defined at the species level. ITS data were almost completely missing from databases. With our work we thus open ways for eventual ITS-based molecular discrimination between species and strains of the *Synechocystis* genus and present data that would be of equal interest for taxonomists, ecologists and evolution biologists investigating unicellular cyanobacteria.

# 2. Experimental

## 2. 1. Cyanobacterial Strains

Cyanobacterial strains used in our study are listed in Table 1. They were all obtained from established culture collections specialized in maintaining microalgae, except for *S. nigrescens* that was obtained from a general supplier of teaching consumables. Strain collections and their acronyms that appear in strain codes were: Culture Collection of Algae at Goettingen University (SAG), The Culture Collection of Algae and Protozoa (Scotland) (CCAP), Culture Collection of Autotrophic Organisms (Institute of Botany of the Academy of Sciences, Czech Republic) (CCALA), Pasteur Culture Collection of Cyanobacteria (PCC) and Carolina Biological Supply Company (Carolina). Most of the strains were catalogued with species names, except for three that were labelled with the genus and strain name/code. Although most of the strains were listed as non-axenic, microscopic inspection after several weeks of growth in our laboratory showed no or only minor contamination with other microorganisms.

All strains were cultured in liquid BG-11 medium (Sigma-Aldrich) with pH adjusted to 7.5 with 1 M HEPES, pH 8.6 (Calbiochem OmniPur grade) under constant cool white light (intensity of 25 μmol/$m^2$s +/- 15%) and at room temperature (22 – 25°C). *Synechocystis nigrescens* was cultured in buffered BG-11 medium as above, with added NaCl to 500 mM final concentration.

**Table 1**. List of cyanobacterial strains used for morphological and sequence analyses

| Species | Strain |
|---|---|
| *Synechocystis aquatilis* | SAG 90.79 |
| *Synechocystis bourrellyi* | CCAP 1480/1 |
| *Synechocystis fuscopigmentosa* | CCALA 810 |
| *Synechocystis limnetica* | CCAP 1480/5 |
| *Synechocystis minuscula* | SAG 258.80 |
| *Synechocystis nigrescens* | Carolina |
| *Synechocystis pevalekii* | SAG 91.79 |
| *Synechocystis salina* | CCALA 192 |
| *Synechocystis* sp. | CCAP 1480/4 |
| *Synechocystis* sp. | PCC 6714 |
| *Synechocystis* sp. | PCC 6803 |

## 2. 2. Polymerase Chain Reaction

Cells from mid- to late exponential phase culture (1 ml) were pelleted by centrifugation. Supernatant was discarded and cells were resuspended in 40 μl of sterile $dH_2O$ and heated for 10 min at 95°C. The lysed cells were used directly for PCR. Reactions were carried out in 20 μl mixtures containing 1 μl of boiled cell suspension, 1 × *Taq*-buffer with $(NH_4)_2SO_4$, 2.5 mM $MgCl_2$, 100 μM of each dNTP, 0.5 μM of each primer (Table 2) and 0.5 U of *Taq*-polymerase (Thermo Scientific), which was added to reaction mixtures after the initial denaturation. PCR reactions were carried out using the following programme: initial denaturation at 95 °C for 5 min, 30 cycles of 95 °C for 30 s, annealing at 55 °C (for ITS amplification) or 60 °C (for 16S rRNA gene amplification) for 30 s and elongation at 72 °C for 1 min (16S) or 2.5 min (ITS) with a final extension step at 72 °C for 7 min. PCR products were resolved on 1.2% or 1.5% agarose gels and visualized using ethidium bromide.

**Table 2**. List of specific primers used for amplification of 16S rDNA and ITS

| Primer | Region | Primer sequence (5'-3') | Reference |
|---|---|---|---|
| CSIF | ITS | GTC ACG CCC GAA GTC GTT AC | [18] |
| ULR | ITS | CCT CTG TGT GCC TAG GTA TC | [18] |
| CY23R | ITS | CTC ATT CTT CAA CAG GCA C | This study |
| CYA106F | 16S | CGG ACG GGT GAG TAA CGG TGA | [2] |
| CYA781Ra | 16S | GAC TAC TGG GGT ATC TAA TCC CAT T | [2] |
| CYA781Rb | 16S | GAC TAC AGG GGT ATC TAA TCC CTT T | [2] |

We constructed the CY23R primer based on sequence alignment of 23S regions of 24 cyanobacterial species from 16 different genera found in sequence databases. A conserved region, identical in all aligned sequences (5' - GTGCCTGTTGAAGAATGAGCCGGCGA – 3') was used to design a primer with appropriate length and melting temperature to be used with the standard cyanobacteria-specific forward primer CSIF. Schematic representation of all the primers used is shown in Fig 1. CYA781Ra and CYA781Rb were always used as an equimolar mixture (0.5 μM) of both, in combination with 0.5 μM forward primer CYA106F.[2]

corresponds to nucleotide positions 90–751 (spanning variable regions V2-V4) in *Synechocystis* sp. PCC 6803 16S rRNA gene as it has proven to be useful for identification of cyanobacteria.[2] From ITS amplicons, the region spanning conserved domains D1 to D5 was analysed.[26] All the sequences were compared to the non-redundant dataset of the GenBank collection using BLASTN.[27] Individual pairwise alignments between sequences were performed using EMBOSS Water algorithm at the EMBL-EBI web server[28] and multiple sequence alignments using MUSCLE algorithm in MEGA version 6 [29] for ITS regions or concatenated 16S and ITS. For multiple alignments of 16S sequences

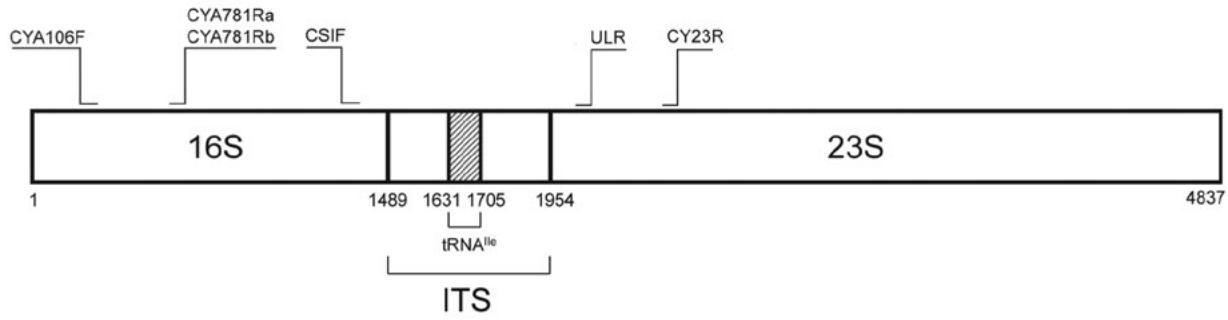


**Figure 1:** Primer positions relative to the 16S and 23S coding regions. Nucleotide positions are labelled for reference as deducted from *Synechocystis* sp. PCC 6803 genome.

## 2. 3. Cloning and Sequencing

After electrophoresis, PCR products were excised from agarose gels and purified using GeneJet Gel Extraction Kit (Thermo Scientific). Purified products were ligated into pJET1.2 using CloneJET™ PCR Cloning Kit (Thermo Scientific). After transformation of competent *Escherichia coli* DH5α cells and plating onto selective media, plasmid DNA was isolated from overnight cultures of one to several independent clones using Plasmid MiniPrep Kit (Thermo Scientific). Sequencing was performed by Macrogen Europe using compatible universal primers annealing to the plasmid backbone.

## 2. 4. Sequence Analyses

For sequence comparisons, only the polymorphic segment of the 16S rRNA gene or the ITS region were used. For 16S rRNA analyses we used the region which

we utilized RDP Aligner.[30] Analyses of tRNA composition in sequenced ITS regions were performed both manually, by finding the conserved segments of ITS in multiple alignments and comparing them to known consensus sequences for $tRNA^{Ile}$ and $tRNA^{Ala}$, as well as with tRNAscan-SE v.1.21 program via the Lowe Lab Webserver Interface.[31]

In addition to cyanobacterial strains listed in Table 1, we investigated in detail the 16S rRNA coding and ITS regions of all three *Synechocystis* sp. strains whose complete genomes were sequenced up to now: PCC 6714, PCC 6803 and PCC 7509. These sequences are available in GenBank under ID codes CP007542.1, CP003265.1 and NZ_ALVU02000001.1, respectively.

Maximum-likelihood trees were built using MEGA version 6 [29] applying the Jukes-Cantor model. Bootstrap resampling using 1000 replicates was performed to test the robustness of the trees. We built 3 trees, based on 16S, ITS or concatenated 16S and ITS sequences using sequences

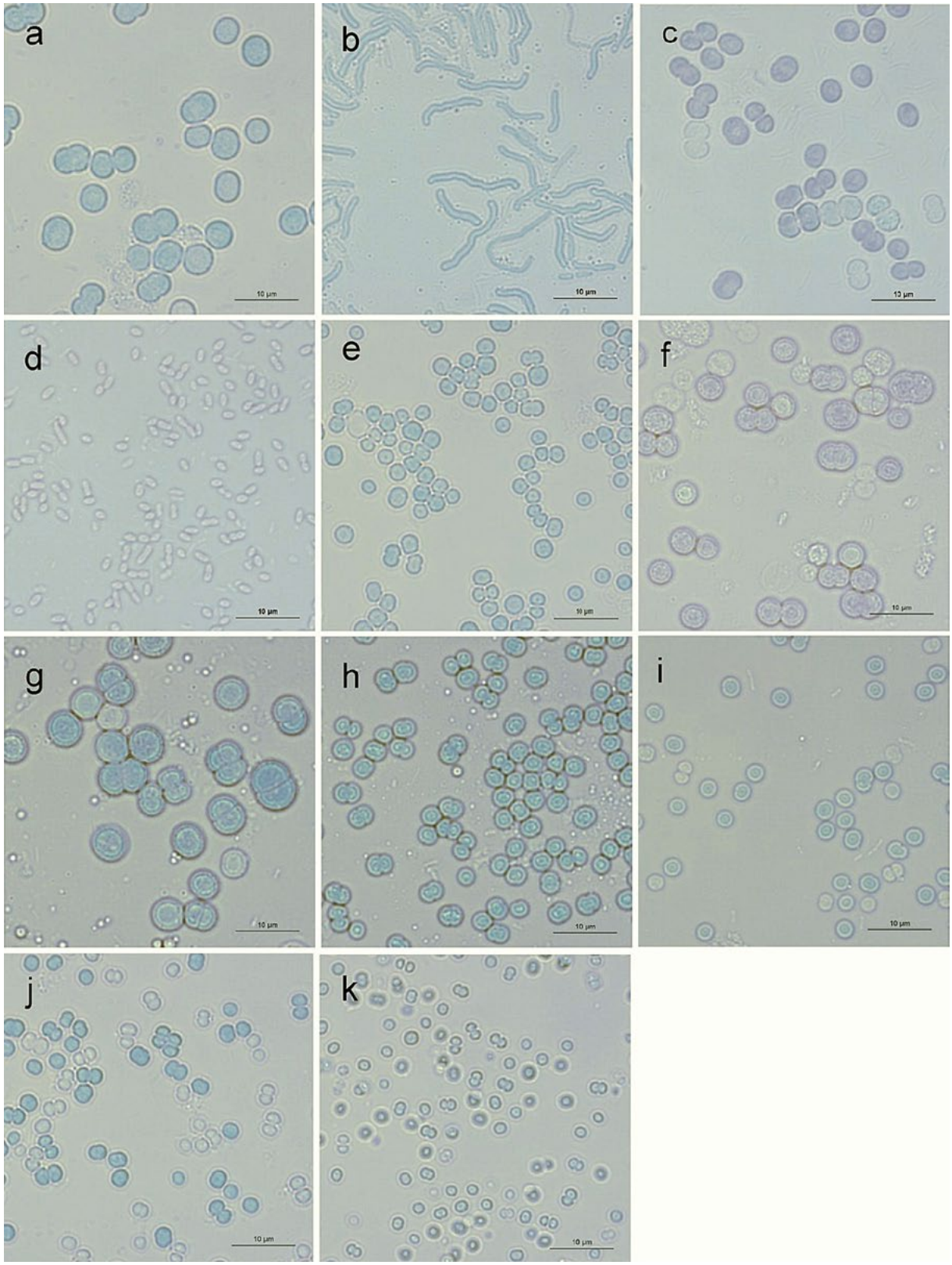


**Figure 2:** Microphotographs of *Synechocystis* strains at 1000× magnification. (a) *Synechocystis aquatilis* SAG 90.79, (b) *Synechocystis bourrellyi** CCAP 1480/1, (c) *Synechocystis fuscopigmentosa* CCALA 810, (d) *Synechocystis limnetica** CCAP 1480/5, (e) *Synechocystis minuscula* SAG 258.80, (f) *Synechocystis nigrescens*, (g) *Synechocystis pevalekii* SAG 91.79, (h) *Synechocystis salina* CCALA 192, (i) *Synechocystis* sp. CCAP 1480/4, (j) *Synechocystis* sp. PCC 6714, (k) *Synechocystis* sp. PCC 6803. Inverted microscope Nikon EclipseTE300 was used. Scale bar corresponds to 10 μm. * denotes species samples with atypical morphology for *Synechocystis* members.

from strains analysed in this study (9 sequences for trees based on ITS and concatenated 16S and ITS sequences, and 10 for the tree based on 16S sequences, since from *Synechocystis nigrescens* we could only amplify 16S rRNA but not ITS region), sequences from two other *Synechocystis* strains with published whole genome sequence (*Synechocystis* sp. PCC 7509 and PCC 6714) and sequences from 4 fully sequenced non-Synechocystis strains, whose 16S or ITS regions showed high similarity to some of our analysed strains.

## 3. Results and Discussion

### 3. 1. Microscopic Investigation of the Strains

Microphotographs of *Synechocystis* strains at 1000× magnification are presented in Fig 2. Cells of strains that later proved to be phylogenetically closest to *Synechocystis* sp. PCC 6803 and PCC 6714 (*Synechocystis salina* CCALA 192, *Synechocystis* sp. CCAP 1480/4 and *Synechocystis minuscula*) were similar in shape and size (1–2.5 (5) μm) to the typical morphology[32] of *Synechocystis* members.

*Synechocystis limnetica* CCAP 1480/5 and *Synechocystis bourrellyi* CCAP 1480/1 resembled shape characteristics of *Synechococcus* genus members. Especially *Synechocystis bourrellyi* with cells several times longer than wide is evidently morphologically different from *Synechocystis* representatives and fits into description of *Synechococcus*-type cell shape: cells 1.5 up to more than 20 μm long and 0.4 to 6 μm wide, according to CyanoDB (http://www.cyanodb.cz/Synechococcus). We thus decided to interpret sequence data obtained with these two strains with care and from here on we label both strains with and asterisk (*) after the species name.

Three of the analysed strains showed cell diameters relatively large for the *Synechocystis* members. *Synechocystis pevalekii* SAG 91.79, *Synechocystis nigrescens* and *Synechocystis aquatilis* SAG 70.79 with diameters ranging from 3.5 to 5 μm represent this group. Although the typical diameter for *Synechocystis aquatilis* is expected to be 4.5 to 7 μm,[33] these cells are larger than typical[32] for *Synechocystis* members. According to their size, these three strains are similar to *Geminocystis* genus members (3–10 μm). Nevertheless we kept these strains for DNA analysis to find out the level of their relatedness to strains with the typical shape and size of *Synechocystis* members.

It has been observed before that cyanobacterial systematics that is based on morphology alone is problematic, as cells change morphology in varying growth conditions. This has for example been shown for the picobacterium *Cyanobacterium aponinum* that displays a very different habitus in salt water (elongated cells) as compared to freshwater.[34] In the literature, there are also reports that growth in laboratory conditions can alter cell phenotype as compared to natural growth conditions.[35] A DNA-based analysis has a clear advantage over microscopic analysis in that it is not affected by eventual changes in cell morphology. On the other hand, with PCR-based methods there is a risk of polymerase errors and cross-contamination, possibly leading to ambiguous results.[36] Furthermore, amplification of DNA from a minor population in non-axenic cultures can occur, especially when broad-specificity primers are used.[2] A microscopic check of the starting ma-

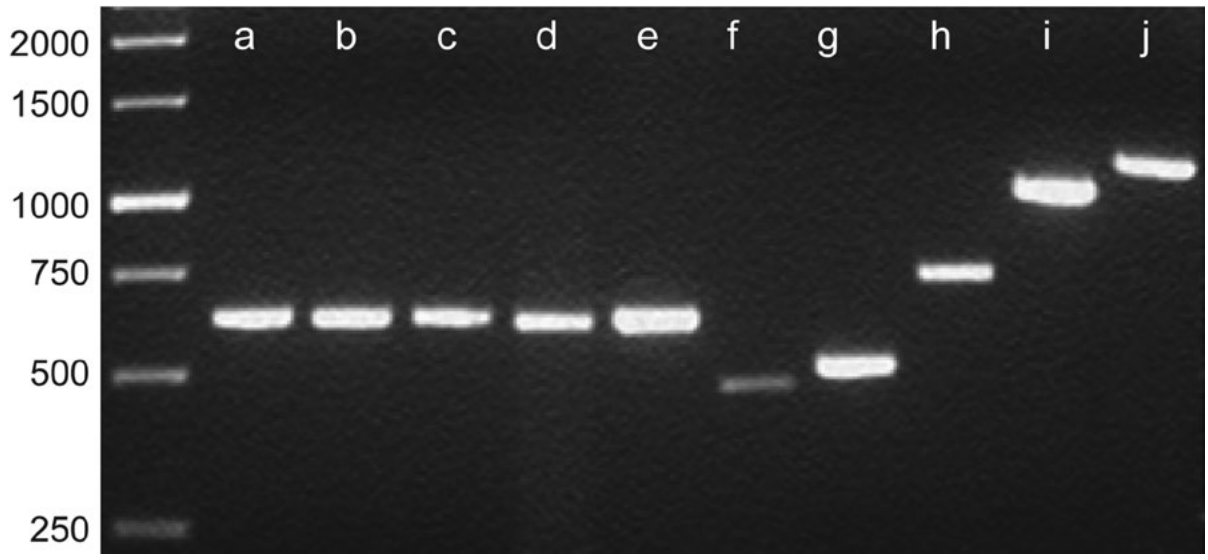


**Figure 3:** PCR amplification products of ITS regions for the 10 *Synechocystis* strains using CSIF and ULR primers, resolved on 1% agarose gel. (a) *Synechocystis* sp. PCC 6803, (b) *Synechocystis* sp. PCC 6714, (c) *Synechocystis* sp. CCAP 1480/4, (d) *Synechocystis salina* CCALA 192, (e) *Synechocystis minuscula* SAG 258.80, (f) *Synechocystis aquatilis* SAG 90.79, (g) *Synechocystis fuscopigmentosa* CCALA 810, (h) *Synechocystis pevalekii* SAG 91.79, (i) *Synechocystis limnetica** CCAP 1480/5, (j) *Synechocystis bourrellyi** CCAP 1480/1. Marker sizes are labelled to the left of the DNA ladder. * denotes species that in microscopic analysis showed atypical morphology for *Synechocystis* members.

terial is thus always recommended. When we did so, we observed that two of the strains display a morphology that is atypical for *Synechocystis* members (consequently marked with an asterisk) and a few other strains had cells larger than typical for *Synechocystis*.

Without very good knowledge and long-standing expertise in microscopic investigation of unicellular cyanobacteria, cell morphologies might be inconclusive about the identity of the investigated species. The discrimination power of DNA is thus much higher and low-cost whole-genome sequencing might open doors to new approaches to strain identification. For the time being, DNA barcoding that is based on selected genomic regions seems to be a reasonable substitute. Even in the future, when polymorphic genomic regions are better understood, DNA barcoding will enable fast identification, possibly even of single cells.

## 3. 2. Amplification and Cloning of Genomic Regions

For cloning and sequencing of ITS regions, PCR products obtained with CSIF and either ULR or CY23R reverse primer were used (Figure 3). Only with *S. aquatilis*, two PCR products were obtained (only the larger product can be clearly seen in Fig. 3) and sequenced that differed in ITS length. All other samples resulted in one PCR product only. Amplicon lengths using CSIF/ULR primers and deduced ITS lengths as obtained by sequencing are given in Table 3.

**Table 3.** Summary of amplicon lengths using the combination of CSIF and ULR primers and deduced ITS lengths obtained for 10 *Synechocystis* representatives

Amplicon lengths were calculated from respective sequences after plasmid cloning of PCR products obtained with CSIF forward and CY23R or ULR reverse primer. ITS lengths correspond to the region spanning conserved domains D1 to D5.[14] *S. nigrescens* ITS region could not be amplified using any of the primers listed in Table 2.

| Species | Amplicon length | ITS length |
|---|---|---|
| *Synechocystis aquatilis* | 418 or 479 | 311 or 312 |
| *Synechocystis bourrellyi** | 1185 | 1018 |
| *Synechocystis fuscopigmentosa* | 512 | 344 |
| *Synechocystis limnetica** | 1056 | 888 |
| *Synechocystis minuscula* | 645 | 477 |
| *Synechocystis pevalekii* | 752 | 587 |
| *Synechocystis salina* | 632 | 466 |
| *Synechocystis* sp. CCAP 1480/4 | 632 | 467 |
| *Synechocystis* sp. PCC 6714 | 631 | 465 |
| *Synechocystis sp*. PCC 6803 | 631 | 465 |

* denotes species that in microscopic analysis showed atypical morphology for *Synechocystis* members.

According to ITS lengths (Table 3), *Synechocystis* members can roughly be divided into four groups. The shortest ITS regions (310-350 bp) were found in *S. aquatilis* and *S. fuscopigmentosa* (group A). Most of the analysed representatives belong to the group B with ITS lengths of between 460 and 480 bp (including *S. minuscula*, *S. salina*, CCAP 1480/4, PCC 6714 and PCC6803). Group C with intermediate size ITS region (587 bp) was represented by *S. pevalekii*, while the eventual group D displayed very long ITS regions (*S. limnetica** 888 bp, *S. bourrellyi** 1018 bp – both were morphologically atypical for *Synechocystis* members as can be seen in Fig. 2). These differences in ITS lengths allow for a rapid PCR-based differentiation between some of the *Synechocystis* members without sequencing, although strain determination cannot be achieved by using universal ITS primers alone.

Iteman et al. reported that ITS regions of cyanobacteria vary in length from 283 to 545 bp,[14] which is with exclusion of *S. bourrellyi** and *S. limnetica** true also for the ITS regions of the analysed *Synechocystis* strains (Table 3). Interestingly, ITS lengths of the two atypical species samples (1018 bp for *S. bourrellyi** and 888 bp for *S. limnetica**) correspond to the lengths that were reported for *Synechococcus* representatives,[37] i.e. between 820 bp (WH 7803) and 1065 bp (PCC 7001). These results are in accordance with the morphological features of the two strains (Fig. 2), displaying characteristics of *Synechococcus* rather than *Synechocystis* species. Taken together, the great variety of the lengths of the ITS segments represents a good starting point for development of amplification-based approaches to differentiation between species and strains within the *Synechocystis* genus.

## 3. 3. Sequence Comparisons

Sequences of 16S rRNA gene variable regions were determined for products of PCR amplification using primers CYA106F and CYA781Ra/b. All the sequences obtained within this work are deposited in GenBank (KT354181– KT354212 and KT371491– KT371499). Respective ID codes are listed in the following sections for each of the strains analysed.

We compared variable segments of 16S rRNA genes and complete ITS sequences from our experiments with those available in GenBank using BLAST. The result of the comparison was a list of sequences with highest levels of identity. Below, we are summarizing our findings for individual species/strains.

In the text, the term 'clone' refers to sequences that we obtained on plasmid-cloned PCR products resulting from amplification of template DNA from individual cyanobacterial cell cultures.

### 3. 3. 1. Synechocystis Aquatilis

*S. aquatilis* is the type species of the genus (Komárek, 2006) and there are several 16S rRNA encoding sequences deposited in the GenBank that enabled their easy align-

ment and analysis of inter-strain differences. We analysed two independent clones of the 16S rRNA region (IDs: KT354181, KT354182). Both our sequences displayed 99.5% identity to the database sequence KM020011.1 originating from the same strain and the same culture collection as ours. Three identical database sequences from 3 *Cyanobacterium aponinum* strains showed the second highest score (97% sequence identity to our sequence): KSU-WH-5 (ID: KT807478.1) collected in Saudi Arabia, lklSCC30 (ID: KM438201.1) collected in Greece and PCC 10605 (ID: CP003947.1), for which the complete genome[38] is available.

Up to now, partial or full 16S rRNA sequences of 8 other *Synechocystis aquatilis* strains have been deposited in GenBank. They did not appear among top-scored hits in our initial sequence comparison and were therefore separately aligned to our sequences using the multisequence alignment program Clustal W2. Comparison of 237 nucleotides shared by all the deposited sequences revealed close relation of our clones to sequences belonging to two different *S. aquatilis* strains (ISB32 and ISB33, isolated from hot springs in Iran) having 99.7% (1 polymorphic site) and 92.5% (18 polymorphic sites) sequence identity, respectively. Sequences of the 16S rRNA from other 6 *Synechocystis aquatilis* strains deposited in GeneBank differed substantially from our newly determined sequences and seem only distantly related to SAG 90.97. Either the strains are genetically substantially polymorphic or the depositors failed to properly determine the species.

BLASTN sequence similarity search comparing our 4 clones of the ITS region positioned *Cyanobacterium aponinum* PCC 10605 as the top match with 95% sequence identity. Except ours, there are no ITS sequences attributed to *Synechocystis aquatilis* currently deposited in GenBank.

### 3. 3. 2. *Synechocystis bourellyi**

Sequences of two 16S rRNA-coding clones (IDs: KT354187, KT354188) and of one ITS region clone (KT354189) were compared to the complete GenBank dataset. The highest score (99.5% identity, 3 mismatches for KT354187 and 99.7% identity or 2 mismatches for KT354188) was shared with various strains of the *Synechococcus* genus *(Synechococcus elongatus* CCAP 1479/1B (ID: KM020008.1), *Synechococcus* sp. CCAP 1479/10 (ID: HE975006.1), *Synechococcus* sp. PCC 7009 (ID: AM709628.1), S*ynechococcus* sp. EW15 (ID: DQ275602.1 and *Synechococcus* sp. BO8806 (ID: AF317072.1)). Comparison with complete genome sequences showed *Cyanobium gracile* PCC 6307 (ID: CP003495.1) with 99.2% (5 mismatches with KT354187) or 99.4% sequence identity (4 mismatches with KT354188) as the highest scoring result.

The ITS region we have amplified was unexpectedly long (Table 3 and Fig. 3). BLASTN search identified *Synechococcus* sp. PCC 7009 (ID: AM709628.1) as the highest score with only two mismatched nucleotides. As with 16S rRNA coding regions, complete genome sequence with the highest score was that of *Cyanobium gracile* PCC 6307 (ID: CP003495.1) with 92.9% sequence identity.

Both 16S rRNA encoding and ITS region sequences thus demonstrate highest identities with members of the *Synechococcus* genus, but also of other related genera. This is in line with the microscopic observations (Fig. 2). *Synechocystis* members did not appear as top scores in the sequence comparisons we have performed.

### 3. 3. 3. *Synechocystis fuscopigmentosa*

Two identical 16S rRNA-coding sequences were obtained (ID: KT371491) displaying 98.9% identity to the corresponding region of *Geminocystis* sp. NIES-3709 (ID: AP014821.1).

Next, we analysed two sequences of the ITS region and also found them identical (ID: KT371492). BLASTN search results showed sequence from *Cyanobacterium* sp. PAP1 (ID: EF555569.1) as the most similar one, but the coverage was not complete since the GenBank submission for PAP1 strain does not contain full ITS sequence. *Geminocystis* sp. NIES-3709 (ID: AP014821.1) displayed the highest overall score among available sequences with complete coverage (96.8% identity).

### 3. 3. 4. *Synechocystis limnetica**

Two 16S rRNA-coding clones were sequenced and analysed (IDs: KT354190, KT354191). Sequence alignment showed that among the cyanobacterial 16S rRNA sequences deposited in databases, *S. limnetica** has the highest similarity with *Synechococcus* sp. MA0607K (ID: FJ763779.1), having 8 or 9 mismatches (for the two clones) in the variable segment alone. Sequence of the ITS region (1 clone sequenced; KT354192) has the highest identity, 87.6%, with *Prochlorococcus marinus* MIT9313 (whole genome, ID: BX548175.1). BLASTN search resulted in sequences with higher identity to our clone (up to 98%), but they were assigned to uncultured and taxonomically undefined organisms. Although *Synechocystis limnetica** is highly related to *Synechocystis bourrellyi** (97%) in the 16S variable region, it differs substantially in the ITS region (57%), as can be seen from Tables 4 and 5.

### 3. 3. 5. *Synechocystis minuscula*

Two clones of the 16S region were identical in sequence (ID: KT354193). The top search result after BLASTN sequence similarity analysis was a GenBank entry KM019989.1 from essentially the same strain, albeit 1 mismatch was detected. The second best results were *Synechocystis salina* LEGE 06155 (ID: HQ832911.1, isolated from the intertidal zone in Northern Portugal) and *Synechocystis* cf. *salina* LEGE 07073 (ID: HM217083.1, isolated

from an estuarine habitat, also in Northern Portugal), both with 97.4% identity.

In the GenBank database we found a 16S rRNA coding sequence of another *Synechocystis minuscula* strain (AICB 62; ID: KJ746516.1), but it displayed only low identity (86.8%) with the sequence of our analysed strain. The AICB 62 strain originated from the Algal and Cyanobacterial Collection (AICB) of the Institute of Biological Research from Cluj-Napoca, Romania.

The 4 clones of the ITS region differed only in the first nucleotide position so that pairs of sequences KT354195/KT354196 and KT354194/KT354197 were identical. They had the highest alignment score with the sequence of *Synechocystis* sp. PAK13 (ID: EF555571.1) and *Synechocystis* sp. PAK12 (ID: EF555570.1)[32] with 84.7% identity, but these PAK strains sequences had only 75% of the total ITS region length covered. The best result with the full coverage of the ITS region was with *Gloeothece* sp. PCC 6909 (CCAP 1480/4, ID: HE975009.1), having 80% identity.

### 3. 3. 5. *Synechocystis nigrescens*

Two 16S rRNA coding sequences were analysed (IDs: KT354198 and KT354199). They displayed one mismatch when compared to each other. BLASTN analysis identified *Synechocystis* sp. SAG 37.92 (ID: KM020010.1) as the highest score with only one mismatch. All other sequences with high similarity were assigned to genera *Geminocystis* or *Synechocystis*.

The ITS region could not be analysed because we were unable to amplify it using any of the primer combinations from Table 2. This might point to the fact that the 5' amplification primer was not hybridizing with the template despite the fact that the annealing region seems to be highly conserved[2] among different cyanobacteria.

### 3. 3. 6. *Synechocystis pevalekii*

Two 16S rRNA-coding sequences were analysed (IDs: KT354200 and KT354201). BLASTN analysis surprisingly showed *Chamaesiphon subglobosus* PCC 7430 (ID: AY170472.1) as the hit with the highest score with only 2 (ID: KT354200) or 3 (ID: KT354201) mismatches in the variable region of the 16S rRNA gene. Comparison with complete genome sequences showed 16S rRNA gene from *Chamaesiphon minutus* PCC 6605 (ID: CP003600.1) as the highest scoring sequence with 97.3% identity.

For the ITS region, we analysed 5 clones (IDs: KT354202 – KT354206). All of them displayed 90% sequence identity with *Chamaesiphon minutus* PCC 6605 (complete genome, ID: CP003600.1). All other hits were less related to the *S. pevalekii* sequence in this region.

Interestingly, microscopic observations of *Synechocystis pevalekii* SAG 70.79 showed almost no morphologic characteristics of the genus *Chamaesiphon* in contrast to our sequence alignment results.

### 3. 3.7. *Synechocystis salina*

Two 16S rRNA-coding sequences were analysed (IDs: KT354209, KT354210). The highest alignment score obtained was that of *Gloeocapsa alpicola* FACH-400 (ID: JX872524.1; three mismatches with KT354209 and one with KT354210) and *Gloeothece* sp. PCC 6909 (CCAP 1480/4, ID: HE975009.1; three mismatches with both clones). *Gloeocapsa alpicola* has been reclassified among genera twice; first it has been assigned to *Synechocystis* genus and lately ordered[32] into a new genus as *Geminocystis herdmanii*. Complete genome sequence with the highest score was that of *Synechocystis* sp. PCC 6803 (ID: CP003265.1) with 98.2% identity in the variable part of the 16S rRNA coding region.

Two sequences of the ITS region (IDs: KT354211, KT354212) were found to contain two polymorphic sites. There was 99.8% (1 mismatch with KT354211) or 100% identity (KT354212) with *Gloeothece* sp. PCC 6909 (CCAP 1480/4, ID: HE975009.1). The next highest identity score was obtained with *Synechocystis* sp. PAK12 (ID: EF555570.1), displaying 93.2% identity.

### 3. 3. 8. *Synechocystis* sp. CCAP 1480/4

It should be noted that in the culture collection, CCAP 1480/4 strain is described as *Synechocystis* sp., while in GenBank this same strain is labelled as *Gloeothece* sp. PCC 6909. Two clones of 16S rRNA coding sequence were analysed. They were identical (ID: KT354207) and showed one mismatch when compared to 16S sequences of both *Gloeocapsa alpicola* FACHB-400 (ID: JX872524.1) and *Gloeothece* sp. PCC 6909 (CCAP1480/4, ID: HE975009.1). The closest complete genome sequence was that of *Synechocystis* sp. PCC 6803 (ID: CP003265.1) with 98.5% identity (10 mismatches across the variable part of the 16S rRNA sequence).

The ITS region (ID: KT354208) displayed 100% identity with *Gloeothece* sp. PCC 6909 (CCAP 1480/4, ID: HE975009.1). The next highest score was that of *Synechocystis* sp. PCC 6714 (complete genome, ID: CP007542.1) having 91.3% identity.

*Gloeothece* members are characterized by formation of small colonies which are enveloped in mucilagous envelopes while *Synechocystis* does not form microcolonies. Our observations (Fig. 2) showed no characteristic envelopes in the strain analysed.

### 3. 3. 9. *Synechocystis* sp. PCC 6714

The genomic sequence of *Synechocystis* sp. PCC 6714 has previously been determined,[39] therefore only one clone of its ITS region (ID: KT371499) was sequenced. It showed 3 mismatches to the genomic sequence of this strain deposited in GenBank (ID: CP007542.1).

### 3. 3.10. *Synechocystis* sp. PCC 6803

Essentially, results with the *Synechocystis* sp. PCC 6803 strain were as expected from the genomic sequence,[40]

although 4 polymorphic sites were found in 16S rRNA coding sequences in our 4 clones (IDs: KT371493 – KT371496). None of our clones was identical to any other published sequence and all 4 had *Synechocystis* sp. LMECYA 68, a strain from Cyanobacteria Culture Collection Estela Sousa e Silva in Portugal (ID: EU078508.1), as the highest BLASTN hit, followed by three *Synechocystis* sp. strains: PUPCCC 62 (ID: KF475890.1), an isolate from India, and KSU-AQIQ-1 (ID: LN997853.1) and KSU-WH-2 (ID: KT807477.1), both discovered in Saudi Arabia. Sequences of these three strains were identical to that in the deposited genomic sequence of *Synechocystis* sp. PCC 6803. Sequence identities for LMECYA 68 strain ranged from 100% with one of our clones to 99.7% (2 mismatches) with another one. In the other three strains (and equally in the published PCC 6803 strain), sequences differed in 1 to 3 positions from our sequence.

Analysis of two clones of ITS sequences (IDs: KT371497, KT371498) showed 0 and 1 mismatches, respectively with the published genomic sequence[40] of *Synechocystis* sp. PCC 6803.

Although *Synechocystis* sp. PCC 6803 is of utmost importance for research on photosynthesis, evolution, as well as for biotechnology and synthetic biology, this strain has never been taxonomically defined to the species level. Especially for environmental and biosafety investigations, it would be helpful to assign a species to this strain as well. From our sequence data, the PCC6803 strain is closely related to *Synechocystis salina*, but not identical. Our results show that PCC 6803 is a distinct taxonomic entity despite the fact that it was described as 'corresponding to *S. aquatilis*'[32] based mainly on its morphologic similarity to the type strain. We found out that the ITS regions of these two *Synechocystis* members are very different, sharing only 52% of the sequence, and that also the 16S rRNA coding variable regions are only 86% identical.

A summary of our findings is presented in Tables 4 and 5, showing identities among the variable segment of the 16S rRNA genes and the ITS sequences, respectively, for 11 *Synechocystis* species/strains (10 for the ITS region). Also included in the tables is *Synechocystis* sp. PCC 7509, the only strain with whole genome sequence available besides PCC 6803 and PCC 6714, both of which we analysed independently.

As evident from Table 4, there are three species that in their 16S rRNA gene sequences differ substantially from the remaining *Synechocystis* members in our study, namely *S. bourrellyi** (which only shows substantial similarity with *S. limnetica**), *S. nigrescens* (more closely related only to *S. aquatilis*) and *S. limnetica** (similar only to *S. bourrellyi**).

**Table 4.** Summary of 16S rRNA variable region sequence identities among *Synechocystis* members Sequences used for comparison were obtained in our laboratory, only those of PCC 6714 and PCC 7509 were taken from GenBank. For PCC 6803 our data were in accordance with GenBank sequences. Shade intensity increases with higher values of sequence identity.

| | *S. aquatilis* SAG 90.79 | *S. bourrellyi** CCAP 1480/1 | *S. fuscopigmentosa* CCALA 810 | *S. limnetica** CCAP 1480/5 | *S. minuscula* SAG 258.80 | *S. nigrescens* | *S. pevalekii* SAG 91.79 | *S. salina* CCALA 192 | *Synechocystis* sp. CCAP 1480/4 | *Synechocystis* sp. PCC 6714 | *Synechocystis* sp. PCC 6803 | *Synechocystis* sp. PCC 7509 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| *S. aquatilis* SAG 90.79 | | | | | | | | | | | | |
| *S. bourrellyi** CCAP 1480/1 | 83 | | | | | | | | | | | |
| *S. fuscopigmentosa* CCALA 810 | 95 | 85 | | | | | | | | | | |
| *S. limnetica** CCAP 1480/5 | 83 | 97 | 85 | | | | | | | | | |
| *S. minuscula* SAG 258.80 | 87 | 86 | 88 | 88 | | | | | | | | |
| *S. nigrescens* | 94 | 84 | 87 | 83 | 88 | | | | | | | |
| *S. pevalekii* SAG 91.79 | 88 | 88 | 90 | 88 | 90 | 89 | | | | | | |
| *S. salina* CCALA 192 | 88 | 86 | 89 | 87 | 96 | 89 | 91 | | | | | |
| *Synechocystis* sp. CCAP 1480/4 | 88 | 87 | 89 | 87 | 96 | 90 | 91 | 99 | | | | |
| *Synechocystis* sp. PCC 6714 | 86 | 87 | 88 | 87 | 97 | 87 | 90 | 98 | 98 | | | |
| *Synechocystis* sp. PCC 6803 | 86 | 86 | 88 | 87 | 97 | 87 | 90 | 98 | 98 | 99 | | |
| *Synechocystis* sp. PCC 7509 | 89 | 85 | 90 | 85 | 89 | 89 | 90 | 90 | 90 | 88 | 88 | |

* denotes species that in microscopic analysis showed atypical morphology for *Synechocystis* members

Identities in the ITS region (Table 5) are far lower than in 16S-rRNA coding region and only a few strains clearly converge in a single group, namely PCC 6803, PCC 6714, CCAP 1480/4 and *S. salina*. For other species/strains identity was below 65% when compared to each other within the dataset.

We additionally compared the variable part of the 16S rRNA coding sequences that were determined in our laboratory with those known previously for members of all the major lineages of cyanobacteria (Appendix, Fig. B). *Synechocystis* members from our analysis appear distributed among Chroococcidiopsidales, Chroococcales, Oscillatoriales, and eventually Synechococcales (only the two strains that were evidently different from others by appearance). This is in good accordance with the previously published phylogenetic tree based on 31 protein sequences from all the fully sequenced genomes of cyanobacteria known in 2014 (Fig. 1 in[24]), just that we additionally found *S. pevalekii* as a new member of the genus evading the Chroococcales order, showing relatedness to Oscillatoriales. *S. bourellyi** and *S. limnetica** stand even further apart from the rest of the analysed *Synechocystis* members, further corroborating the idea that they might have been either mislabelled before they came in our laboratory or were incorrectly taxonomically determined at deposition in the culture collection. Another possible explanation would be horizontal gene transfer, since it is known to be common among cyanobacteria, especially for protein-coding genes.[41,42] Further analyses of additional phenotypic and genotypic characteristics would provide unambiguous conclusions about the observed variability.

**Table 5.** Summary of ITS sequence identities among *Synechocystis* members
Sequence data for PCC 6714 and PCC 7509 were taken from GenBank, all other were obtained in our laboratory. For PCC 6803 our sequences data were in accordance with GenBank sequences. Shade intensity increases with higher values of sequence identity.

| | *S. aquatilis* SAG 90.79 | *S. bourrellyi** CCAP 1480/1 | *S. fuscopigmentosa* CCALA 810 | *S. limnetica** CCAP 1480/5 | *S. minuscula* SAG 258.80 | *S. pevalekii* SAG 91.79 | *S. salina* CCALA 192 | *Synechocystis* sp. CCAP 1480/4 | *Synechocystis* sp. PCC 6714 | *Synechocystis* sp. PCC 6803 | *Synechocystis* sp. PCC 7509 |
|---|---|---|---|---|---|---|---|---|---|---|---|
| *S. aquatilis* SAG 90.79 | | | | | | | | | | | |
| *S. bourrellyi** CCAP 1480/1 | 45 | | | | | | | | | | |
| *S. fuscopigmentosa* CCALA 810 | 79 | 47 | | | | | | | | | |
| *S. limnetica** CCAP 1480/5 | 50 | 57 | 49 | | | | | | | | |
| *S. minuscula* SAG 258.80 | 50 | 37 | 55 | 38 | | | | | | | |
| *S. pevalekii* SAG 91.79 | 52 | 42 | 53 | 48 | 48 | | | | | | |
| *S. salina* CCALA 192 | 52 | 36 | 55 | 41 | 78 | 48 | | | | | |
| *Synechocystis* sp. CCAP 1480/4 | 52 | 36 | 55 | 42 | 78 | 49 | 99 | | | | |
| *Synechocystis* sp. PCC 6714 | 53 | 35 | 52 | 43 | 77 | 44 | 91 | 91 | | | |
| *Synechocystis* sp. PCC 6803 | 52 | 35 | 51 | 41 | 77 | 43 | 91 | 91 | 91 | | |
| *Synechocystis* sp. PCC 7509 | 62 | 45 | 61 | 47 | 52 | 63 | 51 | 51 | 51 | 48 | |

* denotes species that in microscopic analysis showed atypical morphology for *Synechocystis* members

Based on our sequence data, we prepared phylogenetic trees based on 16S rRNA coding region (Fig. 4 top), ITS region (Fig. 4 bottom) and concatenated 16S and ITS (see Appendix). Phylogenetic trees that are based on 16S rRNA-coding and ITS sequences alone do not differ substantially from each other. Nevertheless, they do differ slightly in positioning of *S. minuscula* in the cluster closely related to PCC 6803 (but in the 16S rRNA-based tree, its positioning is supported with low bootstrap. Also *Synechocystis* sp. PCC 7509 is positioned differently in 16S rRNA and ITS trees. Although topologies differ slightly, we do not believe that this influences interpretation of our results. Our intention was not to determine definite intrageneric phylogenetic positions of analysed strains but to illustrate that taxonomic positioning of some *Synechocystis* species is not in accordance with their phylogeny even on the genus level. Namely, they show higher sequence relatedness to representatives of genera other than *Synechocystis*, which is evident from both trees, as well as from the tree based on concatenated sequences (see Appendix, Fig. A).

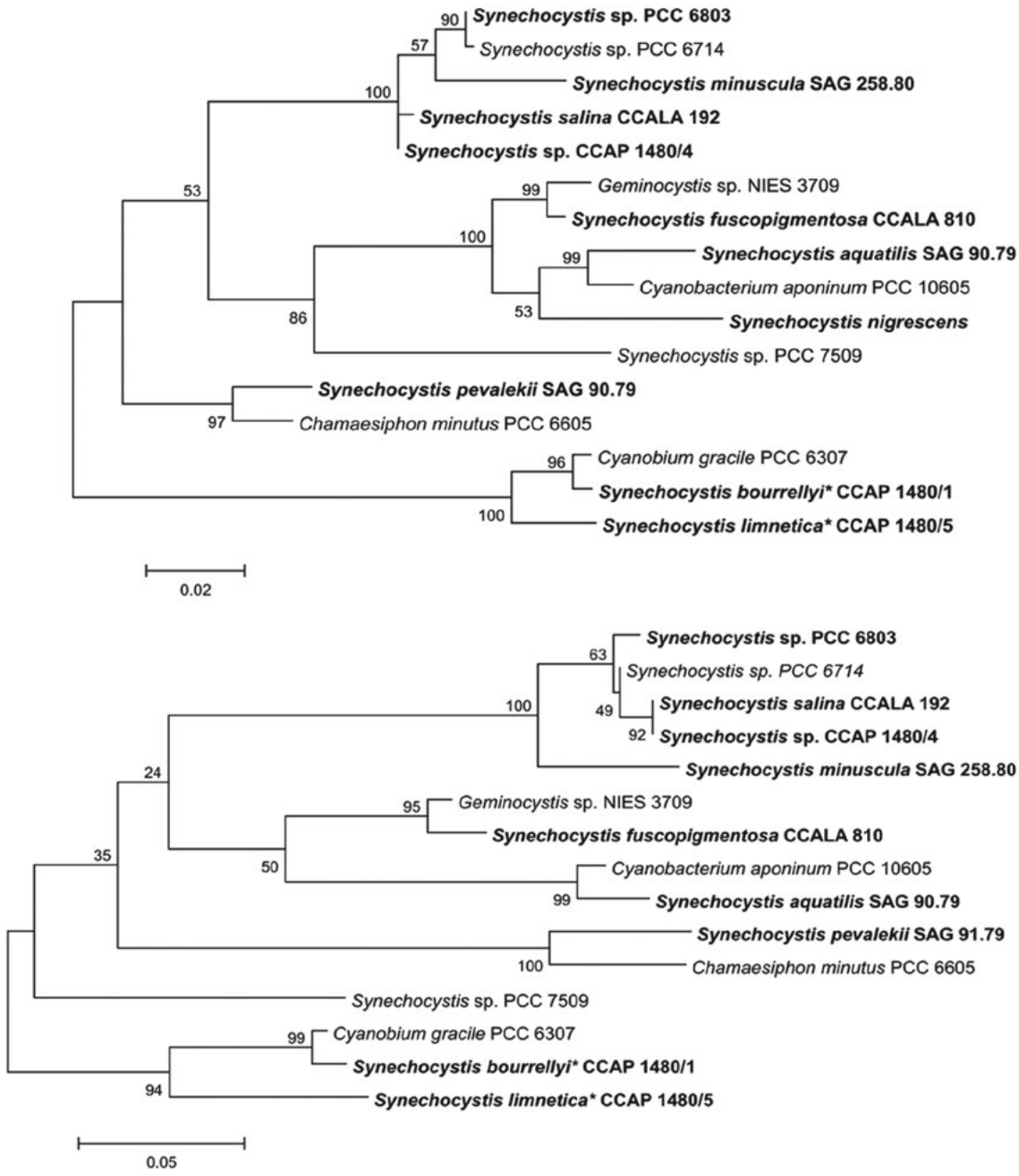


**Figure 4.** Phylogenetic trees for the analysed members of the *Synechocystis* genus based on our sequences (bolded strain names) of the variable segment in 16S rRNA coding region (top) and of ITS regions (bottom). For comparison, sequence data for PCC 6714 and PCC 7509 as deposited in GenBank were included.* denotes strains that in microscopic analysis showed atypical morphology for Synechocystis members.

## 3. 4. tRNA coding Sequences Within ITS Regions

ITS regions in all the strains that we analysed contained $tRNA^{Ile}$ sequences. Only sequences of *S. aquatilis, S. bourrellyi*, S. fuscopigmentosa, S. limnetica** and *S. pevalekii* additionally contained the $tRNA^{Ala}$ sequence, which was not observed in members of the *Synechocystis* genus before. This could be considered an interesting example of the heterogeneity in *Synechocystis*. We found the first case of a two-tRNA ITS in *Synechocystis* in PCC 7509 genome[38] and we further expanded the number of known *Synechocystis* members harbouring 2 tRNA-coding sequences in their ITS to 5 additional species (*S. aquatilis, S. bourrellyi*, S. pevalekii*, *S. fuscopigmentosa* and *S. limnetica**). It remains to be elucidated whether the addition of $tRNA^{Ala}$

coding sequence could have happened through horizontal gene transfer. Alternatively, this could be a sign of a polyphyletic development or, even more likely, of erroneous taxonomic standing of some of the *Synechocystis* species. Again, it cannot be excluded that some strains in culture collections are mislabelled, as e.g. Rajaniemi-Wacklin et al.[43] reported loss of colony structure for cultured *Snowella* strains, upon which they could be easily misidentified as *Synechocystis*. However, *Snowella* (as well as *Woronichinia* and *Merismopedia*) strains from their study were phylogenetically related to *Synechocystis* members.

It has been noted before that more than a half of the strains in the culture collections are probably incorrectly identified.[44] Similarly, Garcia-Pichel et al. discovered that one of the *Microcoleus chthonoplastes* strains in a culture collection and one from a research laboratory were not closely related to fresh isolates and to a cultured strain from another microalgal collection.[45] More recently, Gkelis et al. presented evidence that a *Limnothrix* strain was previously misidentified as a *Planktothrix* strain.[46] DNA-analyses should therefore be used as an important identification factor for culture collections, similarly to what has recently been done[47] on a small scale with a green algae collection from Germany.

Identification of *Synechocystis* and related cyanobacteria in the environmental samples is important from the ecological, but also from the biosafety point of view. *Synechocystis* sp. PCC 6803 is probably the most important cyanobacterial strain in synthetic biology and modern biotechnology. We therefore wished to know whether there are any close relatives of this strain present in aquatic environments and planned to develop a DNA barcoding approach specifically for these unicellular cyanobacteria. In biosafety risk assessments, knowing wild-type relatives of the production strain can help better estimate the risk of e.g. horizontal gene transfer, especially as *Synechocystis* sp. PCC 6803 is known to be naturally competent for transformation.

An extensive review of the current status in cyanobacterial systematics was published by Komárek et al. in 2014.[24] We did not want to go into details of fundamental questions of cyanobacterial taxonomy but instead provide a range of new data that could help in better understanding of the *Synechocystis* genus through its genetic heterogeneity, and eventually contribute to a more precise taxonomic delineation of *Synechocystis* members. In addition, our data could serve as the basis for development of a rapid DNA-based discrimination approach.

The genus *Synechocystis* was listed as one of the polyphyletic genera that need a taxonomic revision.[24] CyanoDB database (http://www.cyanodb.cz/Synechocystis) catalogues as many as 23 *Synechocystis* species described between 1892 and 2006, and three additional species as 'unclear taxa'. Despite our efforts, we could obtain from culture collections around the world only 8 *Synechocystis* representatives that were clearly labelled with a species name. Where several strains of the same species were available, we only analysed one arbitrary chosen strain.

Our search through nucleotide sequence databases revealed that there were relatively few data available for this group of cyanobacteria. Although *Synechocystis* sp. PCC 6803 was the first photosynthetic organism for which a complete genomic sequence was available,[40] there is a considerable gap in understanding genomes of related cyanobacteria. Only two other *Synechocystis* strains were fully sequenced up to now, PCC 6714[39] and PCC 7509.[38] To complement these datasets, there were some sequences of the 16S rRNA-coding regions available for other members of the genus in the sequence databases.

Up to now there has been little work done on comparative genomics of the *Synechocystis* genus. After the first attempt by Korelusová et al.[32] who did the initial comparisons of several strains (not assigned to species) on structural and genetic level, several new sequences were deposited into databases. A report of Kopf et al. focused on a recently sequenced *Synechocystis* sp. PCC 6714 that is closely related to PCC 6803.[48] They showed that the 16S rRNA-coding segment is 99.4% identical to that of PCC 6803, but that almost a quarter of protein-coding genes is unique to each strain.

A recent systematic overview of cyanobacterial genomes encompasses 54 very diverse taxa from across the cyanobacterial phylum that were newly sequenced.[38] Among these, there was the *Synechocystis* sp. strain PCC 7509 that in the phylogenetic tree appeared as only vaguely related to the PCC 6803 strain.

We inspected all three complete genomes of *Synechocystis* genus members for the number and heterogeneity of their rRNA operons. They all contained two identical operons each. In contrast, our sequence analyses show that some strains do display broader heterogeneity in their ITS regions, mostly as single-nucleotide polymorphisms, but also as segment insertions/deletions. Although we did not focus on intrastrain heterogeneity, we provided a clear evidence of ITS polymorphism that is worth considering in developing DNA barcoding tools and elsewhere. It should be noted that cyanobacteria harbour multiple copies of their genome[49] and that there is no clear proof that these copies indeed are identical at the sequence level. Our finding that rRNA sequences are heterogeneous within single strains suggests that 'copies' might differ slightly from each other.

We observed a much greater variability among species in the ITS than 16S rRNA-coding regions, although even 16S rRNA variable sequences differed among several species of the same genus more than we initially expected (Table 4). There were only a few species/strains pairs within the genus that shared >90% identity in the variable segment of the 16S rRNA-coding region. ITS regions were either very similar among strains or quite varied, e.g. *S. minuscula* and *S. pevalekii* display only 48% identity, while *S. salina* and *S. minuscula* share 78% identity in the ITS region (Table 5). This is a good basis for development of

ITS-specific primers that could differentiate between species of the same genus.

Including genomic regions outside the rRNA operon in the analysis could contribute to fine-positioning of genus members into a system, but it was not essential for discrimination between strains, as our results clearly show.

Although our prime interest remains the development of a tool for easy determination of *Synechocystis* members in water bodies, our current results demonstrate the applicability of DNA-based approach in discriminating between species/strains belonging to the same cyanobacterial genus. Moreover, they represent a solid basis for taxonomic reconsideration of *Synechocystis* and related cyanobacterial genera.

## 4. Conclusions

ITS region sequences proved to discriminate among species and strains of *Synechocystis* members and thus represent a solid basis for DNA barcoding. The observed differences between genus members indicate the presence of several genetic clusters which might lead to a taxonomic reinvestigation of the genus. Interestingly, we observed that two out of 11 strains obtained from cell culture collection show morphological and genetic properties different from expected for *Synechocystis* genus members.

Our results greatly expand the range of *Synechocystis* representatives with available genomic sequence data and demonstrate that *Synechocystis* genus currently consists of members that are genetically too different to form one single genus. The need for reconsideration of the genus, previously suggested by Komárek et al.[24] is thus additionally substantiated.

## 5. Acknowledgements

We wish to thank Dr. Bojan Sedmak from the National Institute of Biology for access to the epifluorescence microscope. This project has received funding from the European Union's Seventh Programme for research, technological development and demonstration under grant agreement No 308518, CyanoFactory. Parts of this work have also been supported by the Slovenian Research Agency within the research programme P1-0048a.

## 6. References

1. J. Komárek, *Algae* **2006**, *21*, 349–375. **DOI:**10.4490/ALGAE.2006.21.4.349
2. U. Nübel, F. Garcia-Pichel, G. Muyzer, *Appl. Environ. Microbiol.* **1997**, *63*, 3327–3332.
3. M. Gugger, C. Lyra, P. Henriksen, A. Couté, J. F. Humbert, K. Sivonen, *Int. J. System. Evol. Microbiol.* **2002**, *52*, 1867–1880.
4. N. Engene, R. C. Coates, W. H. Gerwick, *J. Phycol.* **2010**, *46*, 591–601. **DOI:**10.1111/j.1529-8817.2010.00840.x
5. E. M. Eckert, D. Fontaneto, M. Coci, C. Callieri, *Life* **2015**, *5*, 50–64. **DOI:**10.3390/life5010050
6. C. R. Woese, *Microbiol. Rev.* **1987**, *51*, 221–271.
7. S. Smit, J. Widmann, R. Knight, *Nucl. Acids Res.* **2007**, *35*, 3339–3354. **DOI:**10.1093/nar/gkm101
8. N. Engene, H. W. Gerwick, *Fottea* **2011**, *1*, 17–24. **DOI:**10.5507/fot.2011.003
9. G. E. Fox, J. D. Wisotzkey, P. Jurtshuk, Jr., *Int. J. Syst. Evol. Microbiol.* **1992**, *42*, 166–170.
10. N. Walworth, U. Pfreundt, W. C. Nelson, T. Mincer, J. F. Heidelberg, F. Fu, J. B. Waterbury, T. Glavina del Rio, L. Lynne Goodwin, N. C. Kyrpides, M. L. Land, T. Woyke, D. A. Hutchins, W. R. Hess, E. A. Webb, *Proc. Natl. Acad. Sci. U.S.A.* **2015**, *112*, 4251–4256. **DOI:**10.1073/pnas.1422332112
11. D. Honda, A. Yokota, J. Sugiyama, *J. Mol. Evol.* **1999**, *48*, 723–739. **DOI:**10.1007/PL00006517
12. P. S. Seo, A. Yokota, *J. Gen. Appl. Microbiol.* **2003**, *49*, 191–203. **DOI:**10.2323/jgam.49.191
13. V. Piccin-Santos, M. Mendes Brandão, M. Do Carmo Bittencourt-Oliveira, *J. Phycol.* **2014**, *4*, 736–743. **DOI:**10.1111/jpy.12204
14. I. Iteman, R. Rippka, N. Tandeau de Marsac, M. Herdman, *Microbiol.* **2002**, *148*, 481–496. **DOI:**10.1099/00221287-148-2-481
15. G. Rocap, D. L. Distel, J. B. Waterbury, S. W. Chisholm, *Appl. Environ. Microbiol.* **2002**, *68*, 1180–1191. **DOI:**10.1128/AEM.68.3.1180-1191.2002
16. T. A. Sarma. *Handbook of cyanobacteria*. CRC Press, Boca Raton, USA, **2013**. Also available from: http://www.crcnetbase.com/isbn/978-1-4665-5941-7, (accessed January 19, 2017)
17. W. Lu, E. H. Evans, S. M. McColl, V. A. Saunders, *FEMS Microbiol. Lett.* **1997**, *153*, 141–149. **DOI:**10.1111/j.1574-6968.1997.tb10475.x
18. I. Janse, M. Meima, W. E. A. Kardinaal, G. Zwart, *Appl. Environ. Microbiol.* **2003**, *69*, 6634–6643. **DOI:**10.1128/AEM.69.11.6634-6643.2003
19. W. E. A. A. Kardinaal, I. Janse, M. Kamst-van Agterveld, M. Meima, J. Snoek, L. R. Mur, J. Huisman, G. Zwart, P. M. Visser, *Aquat. Microb. Ecol.* **2007**, *48*, 1–12. **DOI:**10.3354/ame048001
20. K. M. Orcutt, U. Rasmussen, E. A. Webb, J. B. Waterbury, K. Gundersen, B. Bergman, *Appl. Environ. Microbiol.* **2002**, *68*, 2236–2245. **DOI:**10.1128/AEM.68.5.2236-2245.2002
21. S. Becker, M. Fahrbach, P. Böger, A. Ernst, *Appl. Environ. Microbiol.* **2002**, *68*, 4486–4494. **DOI:**10.1128/AEM.68.9.4486-4494.2002
22. A. A. Shibl, L. R. Thompson, D. K. Ngugi, U. Stingl, *FEMS Microbiol. Lett.* **2014**, *356*, 118–126. **DOI:**10.1111/1574-6968.12490
23. N. D. Crosbie, M. Pöckl, T. Weisse, *Appl. Environ. Microbiol.* **2003**, *69*, 5716–5721. **DOI:**10.1128/AEM.69.9.5716-5721.2003
24. J. Komárek, J. Kaštovský, J. Mareš, J. R. Johansen, *Preslia* **2014**, *86*, 295–335.

25. J. Komárek, *Hydrobiol.* **2016**, *764*, 259–270. **DOI:**10.1007/s10750-015-2242-0
26. I. Iteman, R. Rippka, N. Tandeau de Marsac, M. Herdman, *Microbiol.* **2000**, *146*, 1275–1286. **DOI:**10.1099/00221287-146-6-1275
27. S. F. Altschul, W. Gish, W. Miller, E. W. Myers, D. J. Lipman, *J. Mol. Biol.* **1990**, *215*, 403–410. **DOI:**10.1016/S0022-2836(05)80360-2
28. W. Li, A. Cowley, M. Uludag, T. Gur, H. McWilliam, S. Squizzato, Y. M. Park, N. Buso, R. Lopez, *Nucl. Acids Res*. **2015**, *43(W1)*, W580–584. **DOI:**10.1093/nar/gkv279
29. K. Tamura, G. Stecher, D. Peterson, A. Filipski, S. Kumar, *Mol. Biol. Evol.* **2013**, *30*, 2725–2729. **DOI:**10.1093/molbev/mst197
30. J. R. Cole, Q. Wang, J. A. Fish, B. Chai, D. M. McGarrell, Y. Sun, C. T. Brown, A. Porras-Alfaro, C. R. Kuske, J. M. Tiedje, *Nucl. Acids Res.* **2014**, *42(D1)*, 633–642. **DOI:**10.1093/nar/gkt1244
31. P. Schattner, A. N. Brooks, T. M. Lowe, *Nucl. Acids Res.* **2005**, *33*, W686–9. **DOI:**10.1093/nar/gki366
32. J. Korelusová, J. Kaštovský, J. Komárek, *J. Phycol.* **2009**, *45*, 928–937. **DOI:**10.1111/j.1529-8817.2009.00701.x
33. J. Komárek, K. Anagnostidis, in: H. Ettl, G. Gärtner, H. Heynig, D. Mollenhauer (Eds.), *Süsswasserflora von Mitteleuropa 19/1*, Gustav Fischer, Jena, Germany, **1998**.
34. I. Moro, N. Rascio, N. La Rocca, M. Di Bella, C. Andreoli, *Algol. Stud.* **2007**, *123*, 1–15. **DOI:**10.1127/1864-1318/2007/0123-0001
35. R. W. Castenholz, J. B. Waterbury, in: J. T. Staley, M. P. Bryant, N. Pfenning, J. G. Holt (Eds.): *Bergey's manual of systematic bacteriology*, Williams and Wilkins, Baltimore, USA, **1989**, pp. 1710–1789.
36. F. von Winztigerode, U. B. Göbel, E. Stackebrandt, *FEMS Microbiol. Rev.* **1997**, *21*, 213–29. **DOI:**10.1111/j.1574-6976.1997.tb00351.x
37. W. Laloui, K. A. Palinska, R. Rippka, F. Partensky, N. Tandeau de Marsac, M. Herdman, I. Iteman, *Microbiol.* **2002**, *148*, 453–465. **DOI:**10.1099/00221287-148-2-453
38. P. M. Shih, D. Wu, A. Latifi, S. D. Axen, D. P. Fewer, E. Talla, A. Calteau, F. Cai, N. Tandeau de Marsac, R. Rippka, M. Herdman, K. Sivonen, T. Coursin, T. Laurent, L. Goodwin, M. Nolan, K. W. Davenport, C. S. Han, E. M. Rubin, J. A. Eisen, T. Woyke, M. Gugger, C. A. Kerfeld, *Proc. Natl. Acad. Sci. U.S.A.* **2013**, *110*, 1053–1058. **DOI:**10.1073/pnas.1217107110
39. M. Kopf, S. Klähn, B. Voss, K. Stuber, B. Huettel, R. Reinhardt, W. R. Hess. *Genome Announc.* **2014**, *2*, e00757–14. **DOI:**10.1128/genomeA.00757-14
40. T. Kaneko, S. Sato, H. Kotani, A. Tanaka, E. Asamizu, Y. Nakamura, N. Miyajima, M. Hirosawa, M. Sugiura, S. Sasamoto, T. Kimura, T. Hosouchi, A. Matsuno, A. Muraki, N. Nakazaki, K. Naruo, S. Okumura, S. Shimpo, C. Takeuchi, T. Wada, A. Watanabe, M. Yamada, M. Yasuda, S. Tabata, *DNA Res.* **1996**, *3*,109–136. **DOI:**10.1093/dnares/3.3.109
41. O. Zhaxybayeva, J. P. Gogarten, R. L. Charlebois, W. F. Doolittle, R. T. Papke, *Genome Res.* **2006**, *16*, 1099–1108. **DOI:**10.1101/gr.5322306
42. S. Yerrapraganda, J. L. Siefert, G. E. Fox, *Methods Mol. Biol.* **2009**, *532*, 339–366 **DOI:**10.1007/978-1-60327-853-9_20
43. P. Rajaniemi-Wacklin, A. Rantala, M. A. Mugnai, S. Turicchia, S. Ventura, J. Komárková, L. Lepistö, K. Sivonen, *J. Phycol.* **2006,** *42,* 226–232. **DOI:**10.1111/j.1529-8817.2006.00179.x
44. J. Komárek, K. Anagnostidis, *Algol. Stud.* **1989**, *56*, 247–345.
45. F. Garcia-Pichel, L. Prufert-Bebout, G. Muyzer, *Appl. Environ. Microbiol.* **1996**, *62*, 3284–3291.
46. S. Gkelis, P. Rajaniemi, E. Vardaka, M. Moustaka-Gouni, T. Tanaras, K. Sivonen, *Microb. Ecol.* **2005**, *49*, 176–182. **DOI:**10.1007/s00248-003-2030-7
47. R. Hoshina, *BMC Res. Notes* **2014**, *7*, 592. **DOI:**10.1186/1756-0500-7-592
48. M. Kopf, S. Klähn, N. Pade, C. Weingärtner, M. Hagemann, B. Voss, W. R. Hess, *DNA Res.* **2014**, *21*, 255–266. **DOI:**10.1093/dnares/dst055
49. K. Zerulla, K. Ludt, J. Soppa, *Microbiol.* **2016**, *162*, 730–739. **DOI:**10.1099/mic.0.000264

## Povzetek

Cianobakterije so pomembna skupina mikroorganizmov z zelo raznoliko morfologijo, na podlagi katere lahko fenotipsko razlikujemo med taksonomskimi linijami cianobakterij. Vendar je morfološko razlikovanje zanesljivo predvsem na ravni rodu, na ravni vrste ali seva pa pogosto ne. Osredotočili smo se na rod enoceličnih cianobakterij *Synechocystis*, ki vključuje tudi modelni cianobakterijski sev PCC 6803. Določili smo zaporedja variabilnega dela genomske regije rRNA 16 S in regije ITS med zapisoma za rRNA 16 S in 23 S za 11 predstavnikov rodu *Synechocystis*. Zaporedja dveh od enajstih analiziranih sevov iz zbirk kultur so se pomembno razlikovala od zaporedij tipičnih predstavnikov rodu *Synechocystis*. Opaženo razlikovanje na molekularni ravni smo potrdili tudi z mikroskopijo. Za ostale seve smo ugotovili, da sta obe genomski regiji, ki se sicer že uporabljata v taksonomiji bakterij, ustrezni za razlikovanje med analiziranimi vrstami, pri čemer regija ITS omogoča tudi zanesljivo razlikovanje med sevi iz rodu *Synechocystis*.