Доклади на Българската академия на науките
Comptes rendus de l'Académie bulgare des Sciences

Tome 70, No 8, 2017

BIOLOGIE

Biologie moléculaire

UTILITY OF THE 5S rRNA SEQUENCE AND ITS SECONDARY STRUCTURE FOR PHYLOGENETIC ANALYSES AND RECOGNITION OF CYANOBACTERIAL STRAINS

Dzhemal Moten, Diyana Basheva*, Rumen Mladenov*, Balik Dzhambazov, Ivanka Teneva*

(*Submitted by Corresponding Member R. Pankov on November 19, 2016*)

Abstract

Cyanobacteria comprise oxygenic phototrophic prokaryotes characterized with high diversity of their morphological, physiological and ecological parameters. The enormous diversity makes their classification complicated and currently the polyphasic approach is most applicable. Suitable molecular markers for resolving species or generic level relationships within Cyanobacteria are still needed. In this study, we have performed phylogenetic analyses based on 5S rRNA gene sequences of 96 cyanobacterial strains by using maximum-likelihood (ML) and neighbour-joining (NJ) algorithms. For comparison, we have used the same strains and methods to generate phylogenetic trees based on the 16S rRNA gene sequences. Both 5S rRNA and 16S rRNA sequences were retrieved from the available cyanobacterial genomes. Our results showed that 5S rRNA could be used as an additional molecular marker in resolving different phylogenetic clades within Cyanobacteria at generic level and above. In addition, we have analyzed the differences between the secondary structures of studied 5S rRNAs. Specific positions of the secondary structure of 5S rRNA could be also used as markers to distinguish cyanobacterial strains at generic level. Thus, the 5S rRNA and its secondary structure are suitable additional markers for solving the taxonomic position and phylogenetic analyses of cyanobacterial strains.

Key words: Cyanobacteria, 5S rRNA, sequences, secondary structure, polyphasic taxonomy

Introduction. The phylum Cyanobacteria (Cyanoprokaryota) is one of the most ancient groups of organisms and includes oxygenic photosynthetic prokaryotes with highly diverse morphology, physiology, ecology and developmental characteristics [1–4]. Since these organisms are considered as progenitors of the chloroplasts in all photosynthetic eukaryotes and some of them (heterocystous cyanobacteria) are able to fix nitrogen, they play a key role in the evolution of the plant kingdom and in the global carbon and nitrogen cycles.

The evolutionary relationships and taxonomy of the cyanoprokaryotes are very complicated and still poorly understood due to their enormous variability, plasticity and horizontal gene transfer. So far, for classification of Cyanobacteria have been used several systems [5], but most accepted is the polyphasic taxonomy, which combines morphological with ecophysiological and molecular characters [2,6]. Recently, Komárek et al. [5] suggested a new phylogenetically-based taxonomic system for Cyanobacteria, in which at least the genera and species should be monophyletic.

At present, the 16S rRNA gene sequence is one of the most widely used molecular markers, which provides the current basis for the phylogeny and taxonomy of Cyanobacteria [7]. In addition, other loci (16S-23S ITS, rpoC1, gyrB, rpoB, rbcLX, cpcBA-IGS) have also been recommended for identification of cyanobacterial species within the polyphyletic genera [8–12].

During the last years, the number of sequenced cyanobacterial genomes increased significantly. The availability of such genome sequences for large number of cyanobacterial strains provides great opportunities for phylogenetic and comparative genomic analyses, which could help to understand the evolutionary relationships within the phylum as well as for the taxonomic classification of these organisms. Based on the genome data, several phylogenetic trees have been generated aiming to find unique conserved signature proteins that could be used to distinguish and clarify the main clades of Cyanobacteria [13–17]. Large sets of proteins that are specific for different phylogenetically defined clades have been reported, but unfortunately, at this stage most of them are still hypothetical [16]. Hence, it is of crucial importance to find suitable and well-defined molecular markers that may be useful for resolving relationships within Cyanobacteria.

So far, the 5S ribosomal RNA gene is not exploited for phylogenetic analyses of Cyanobacteria. It codes a small (approximately 120 nucleotides) ribosomal RNA molecule, which is a structural and functional component of the large ribosomal subunit. 5S rRNA is highly conserved throughout nature. In prokaryotes, it is synthesized as part of a single long transcript, together with 16S and 23S rRNAs and interacts with the ribosomal proteins L5, L18 and L25 [18]. Therefore, in this study, we have used 5S rRNA sequences retrieved from the available cyanobacterial genome sequences to construct phylogenetic trees and secondary structures in order to see whether 5S rRNA could be used as an additional molecular marker in resolving different phylogenetic clades within the phylum Cyanobacteria.

Materials and methods. Ninety-six nucleotide sequences of the 5S rRNA and 16S rRNA regions were chosen after evaluation of the available complete cyanobacterial genomes in the NCBI genome database (`https://www.ncbi.nlm.nih.gov/`). The accession numbers of the used sequences are given in the phylogenetic trees before the names of the cyanobacterial strains.

Analyzed nucleotide sequences of 5S rRNA and 16S rRNA genes were obtained from the cyanobacterial genomes and multiple sequence alignment was performed using the ClustalW tool within alignment function of the software MEGA 7 (`http://www.megasoftware.net/`). Phylogenetic trees were generated by MEGA 7 using the maximum-likelihood (ML), and neighbour-joining (NJ) algorithms [19]. Both algorithms were performed with 1000 bootstrap replicates. Nucleotide positions containing gaps and missing data were eliminated (complete deletion option). The evolutionary distances were computed using the maximum composite likelihood method. For ML analysis, the general time reversible (GTR) model with corrected invariable sites (I), gamma distribution shape parameters (G), and nearest-neighbour-interchange algorithm were selected. The sequences of 5S rRNA and 16S rRNA genes from *Escherichia coli* were used to root the trees.

The secondary structures of the 5S rRNA were generated by using 5S rRNA Database [20] (`http://www.combio.pl/rrna/`) and analyzed either by LocARNA (`http://rna.informatik.uni-freiburg.de/`) or manually.

Results and discussion. In order to evaluate the applicability of the 5S rRNA as an additional molecular marker for phylogenetic analyses of Cyanobacteria, we have constructed phylogenetic trees for 96 cyanobacterial strains based on 5S rRNA and 16S rRNA retrieved from the available whole genome cyanobacterial sequences by using maximum-likelihood (ML), and neighbour-joining (NJ) algorithms. Since the branching patterns of the ML and NJ trees were very similar, here are presented only the ML trees.

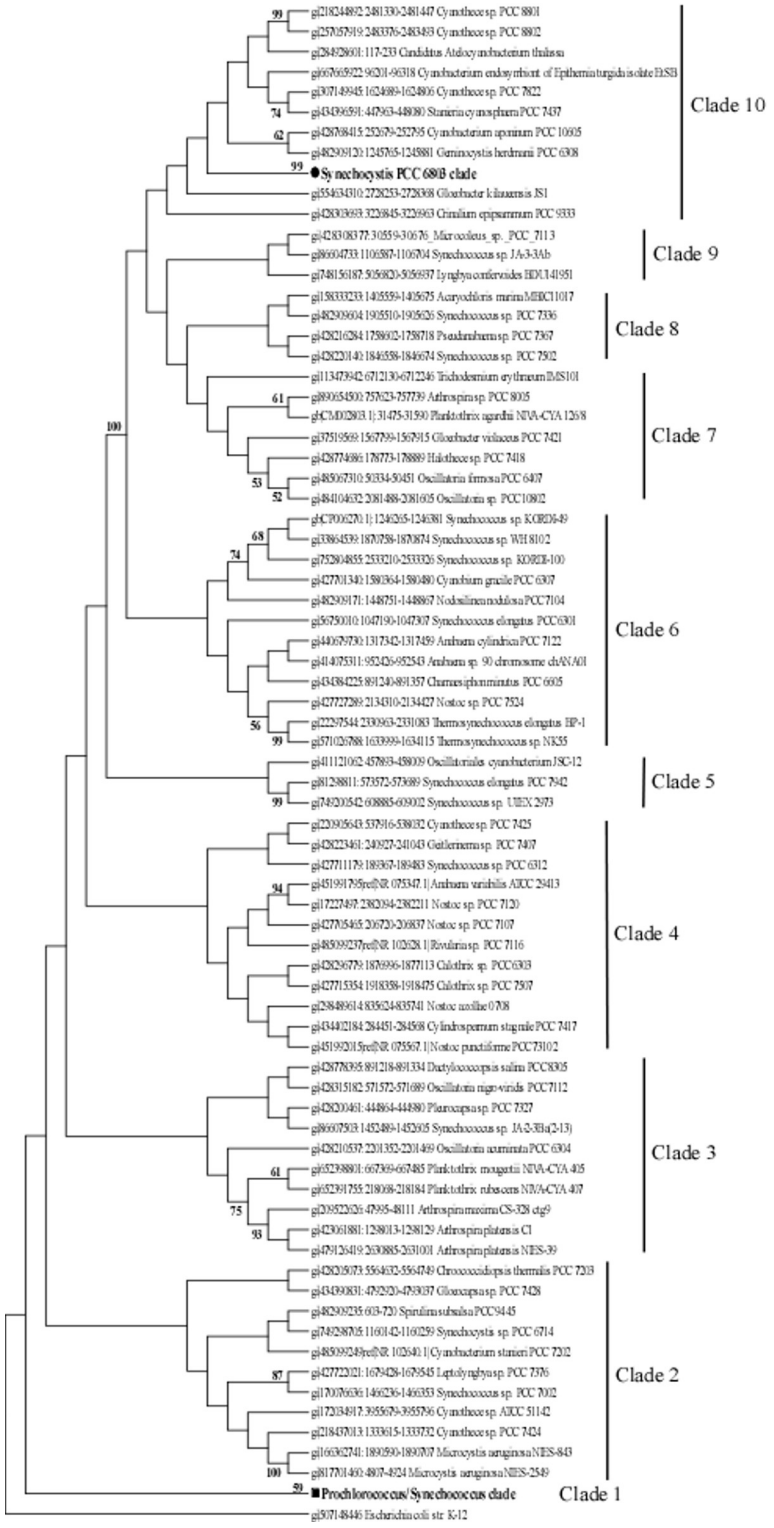
Figure 1 shows a rooted ML distance tree based on the 5S rRNA gene sequences. As can be seen there, cyanobacterial strains are grouped in ten larger clades. Clade 1 (designated as *Prochlorococcus*/*Synechococcus* clade) includes 13 *Prochlorococcus* and 6 *Synechococcus* strains. The subtree that forms Clade 1 was replaced with this sign to simplify the tree and to decrease the displayed strains in the figure since it is comprised only of several *Prochlorococcus marinus* strains and *Synechococcus* strains supported by a bootstrap value of 59%. Other main clades (Clade 2 and Clade 10) are composed of diverse cyanobacterial genera including *Microcystis*, *Cyanothece*, *Synechocystis*, *Cyanobacterium*, *Leptolyngbya*, *Spirulina*, *Stanieria*, etc., which belong to different orders (Chroococcales, Oscillatoriales, Synechococcales, Pleurocapsales). Within these clades only few strains are supported by high bootstrap values including two *Microcystis aeruginosa* strains, two *Cyanothece* strains and five strains *Synechocystis* PCC 6803 designed as *Synechocystis* PCC 6803 clade (Fig. 1). Representatives of Oscil-

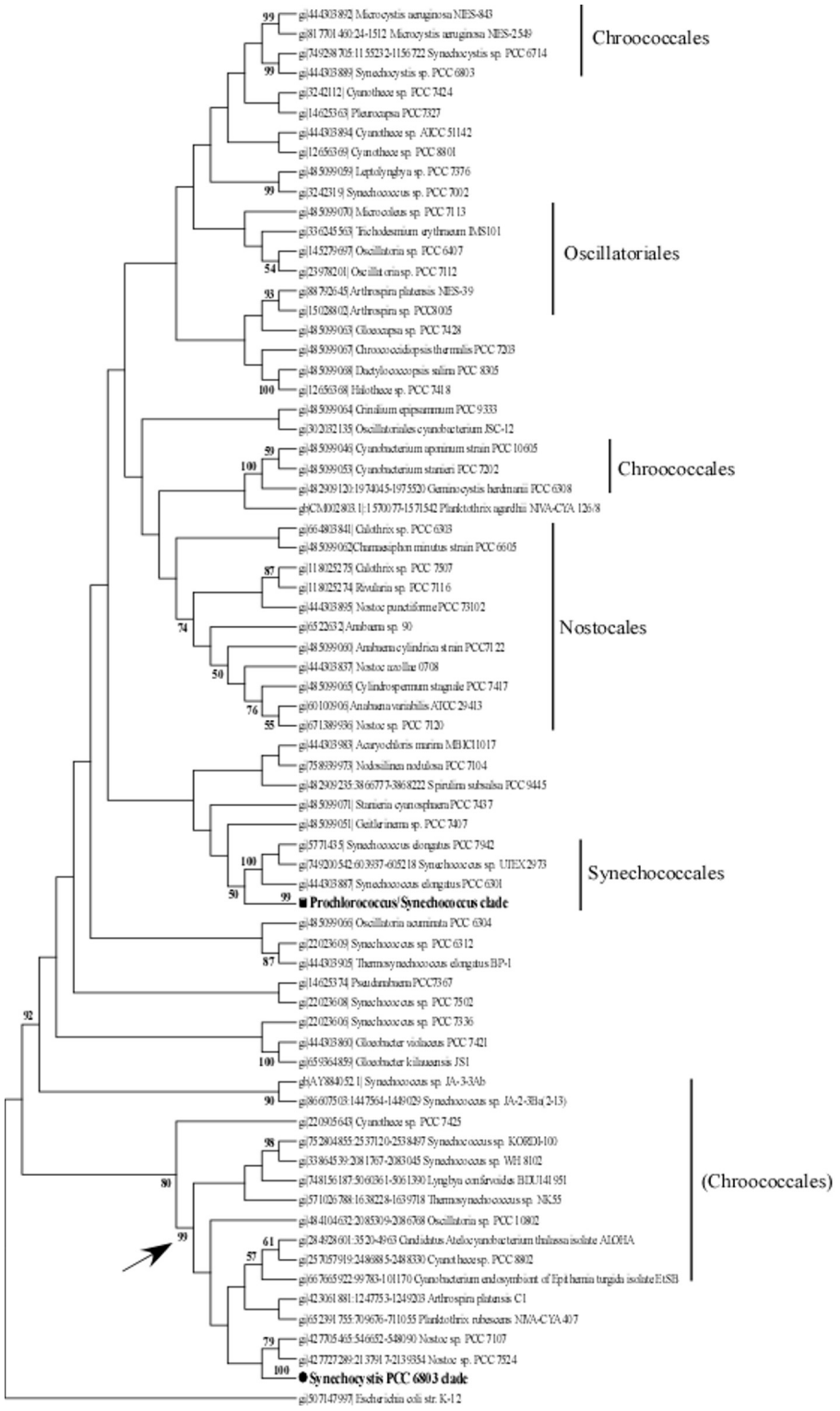
latoriales are clustered in two separate clades (Clades 3 and 7). Usually these filamentous cyanobacteria are most problematic from a taxonomic point of view, because many of the genera are polyphyletic. Based on the polyphasic taxonomy Komárek et al. [5] performed reorganization of this order as some genera were transferred to other orders or new genera were formed. Clade 4 comprises strains that belong to order Nostocales (Fig. 1). Synechococcales strains are grouped in Clade 5 and Clade 6 intermixed with some Nostocales. The rest small clades (Clades 8 and 9) include different cyanobacteria (Fig. 1), which probably need morphological revision as well.

For comparison, in Figure 2 is shown a rooted ML distance tree based on the 16S rRNA for the same cyanobacterial strains as in Fig. 1. At the genus level and above, the sequence of 16S rRNA has been defined as "backbone of prokaryote taxonomy" [21]. This is the reason to compare both 5S rRNA and 16S rRNA trees. There is similarity between these two trees, but they are not equal (Fig. 1, 2). The similarity is related to the clades *Prochlorococcus*/*Synechococcus* (99% bootstrap support) and *Synechocystis* PCC 6803 (100% bootstrap support), which include substrains of these cyanobacteria as well as for the Nostocales (Clade 4) that are clustered in a separate clade supported by a bootstrap value of 74% (Fig. 2). Some of the other cyanobacterial species are also grouped in smaller separate clades representing the orders Oscillatoriales, Synechococcales or Chroococcales (Fig. 2), but not as they are clustered in the 5S rRNA tree (Fig. 1). Comparing the topology of 5S rRNA and 16S rRNA trees, it can be seen that the clades within the 5S rRNA tree are clustered much better than within the 16S rRNA tree. It is noteworthy that within the 16S rRNA tree different cyanobacterial strains that belong to different orders (Synechococcales, Oscillatoriales, Chroococcales) are grouped in one clade supported by a bootstrap value of 99% (Fig. 2, indicated by arrow). If we take into account the older classification of Cyanobacteria (before the revision performed by Komárek et al. [5]), most of these species belong to order Chroococcales. This puts in question whether the classification suggested by Komárek et al. [5] is appropriate. Otherwise, closely related strains from one genus/species (e.g., *Microcystis aeruginosa*, *Synechocystis sp.*, *Synechococcus sp.* or *Arthrospira sp.*) are always grouped together supported by high bootstrap values in both 5S rRNA and 16S rRNA trees (Fig. 1, 2).

The problem with most available cyanobacterial genomes is that many of the sequenced strains are not identified at species level and they are randomly selected

Fig. 1. Phylogenetic analysis by Maximum-Likelihood (ML) method based on 5S rRNA gene sequences. The tree was generated by applying the GTR+I+G evolutionary model. The numbers above branches indicate bootstrap support (> 50%) from 1000 replicates. The tree was rooted using the sequence for *Escherichia coli* K-12. Major clades resolved in the tree are indicated. GenBank accession numbers are given before the species name. The analysis involved 96 nucleotide sequences →


Clade 10
Clade 9
Clade 8
Clade 7
Clade 6
Clade 5
Clade 4
Clade 3
Clade 2
Clade 1
Synechocystis PCC 6803 clade
Prochlorococcus/Synechococcus clade


Chroococcales
Oscillatoriales
Chroococcales
Nostocales
Synechococcales
Prochlorococcus/Synechococcus clade
(Chroococcales)
Synechocystis PCC 6803 clade

based on different criteria (economical importance, availability as culture/isolates, environmental impact, etc.). For the appropriate classification and systematics of Cyanobacteria it is of crucial importance to be sequenced the genomes of generic types represented the main genera, which can be used as reference sequences for the polyphasic taxonomy. Unfortunately, nowadays only few genome projects are based and focused on this approach. With increased number of available whole genome sequences for generic type species of Cyanobacteria, the problems with their taxonomic position and evolutionary relationships will be solved.

In addition, we have analyzed the secondary structures of studied 5S rRNAs. In Figure 3 5S structures of representative cyanobacterial strains from the main clades that are typical for the certain clade are shown. The generalized cyanobacterial 5S RNA structure generated by the software contains 155 nucleotides (Fig. 3). By analyzing several secondary structures of *Prochlorococcus* strains (Clade 1, Fig. 1) we found that position U_{20} (loop A), pairs A_{31}-U_{84} and U_{32}-A_{83} (helix II), U_{60} (loop C) and the pair U_{101}-A_{122} (helix IV) are specific for genus *Prochlorococcus* (Fig. 3A) as most of them (with exception of the pair A_{31}-U_{84}) are shared with genus *Synechococcus* (Clade 1 and Clade 6, Fig. 1). Also, in addition to those signature markers, genus *Synechococcus* has a specific position A_{112} within loop D (Fig. 3E). The genera *Microcystis* (Fig. 3B, Clade 2, Fig. 1) and *Synechocystis* (Fig. 3G, Clade 10, Fig. 1) have common specific nucleotides within loop A (U_{23}), helix II (pairs U_{31}-A_{84} and A_{33}-U_{82}) and helix IV (U_{104}-A_{119}). *Synechocystis* strains have one more specific position (C_{63}) within loop C (Fig. 3G). Although *Arthrospira platensis* (Clade 3, Fig. 1) and *Oscillatoria formosa* (Clade 7, Fig. 1) are clustered in two different clades (Fig. 1), both belong to Oscillatoriales and have similar secondary structures of 5S rRNA (Fig. 3C, F). Their common specific positions are A_{44}, A_{71} (within helix III) and the pair A_{96}-A_{127} (within helix V). Positions C_{112} and A_{113} within loop D are specific only for *Oscillatoria* (Fig. 3F). Cyanobacteria from genus *Nostoc* (Clade 4, Fig. 1) have specificity at positions A_{71} and G_{72} (Fig. 3D), while the *Cyanothece* strains (Clade 10, Fig. 1) have specific nucleotides within loop B (C_{37}), helix III (C_{71} and A_{72}) and helix IV (the pair U_{105}-A_{118}) (Fig. 3H).

These data showed that specific positions of the secondary structure of 5S rRNA could be used as additional markers to distinguish cyanobacterial strains at generic level and above.

Fig. 2. A Maximum-Likelihood (ML) tree based on 16S rRNA gene sequences. The tree was generated by applying the GTR+I+G evolutionary model. The numbers above branches indicate bootstrap support ($> 50\%$) from 1000 replicates. The tree was rooted using the sequence for *Escherichia coli* K-12. Major clades resolved in the tree are indicated. GenBank accession numbers are given before the species name. The analysis involved 96 nucleotide sequences

←

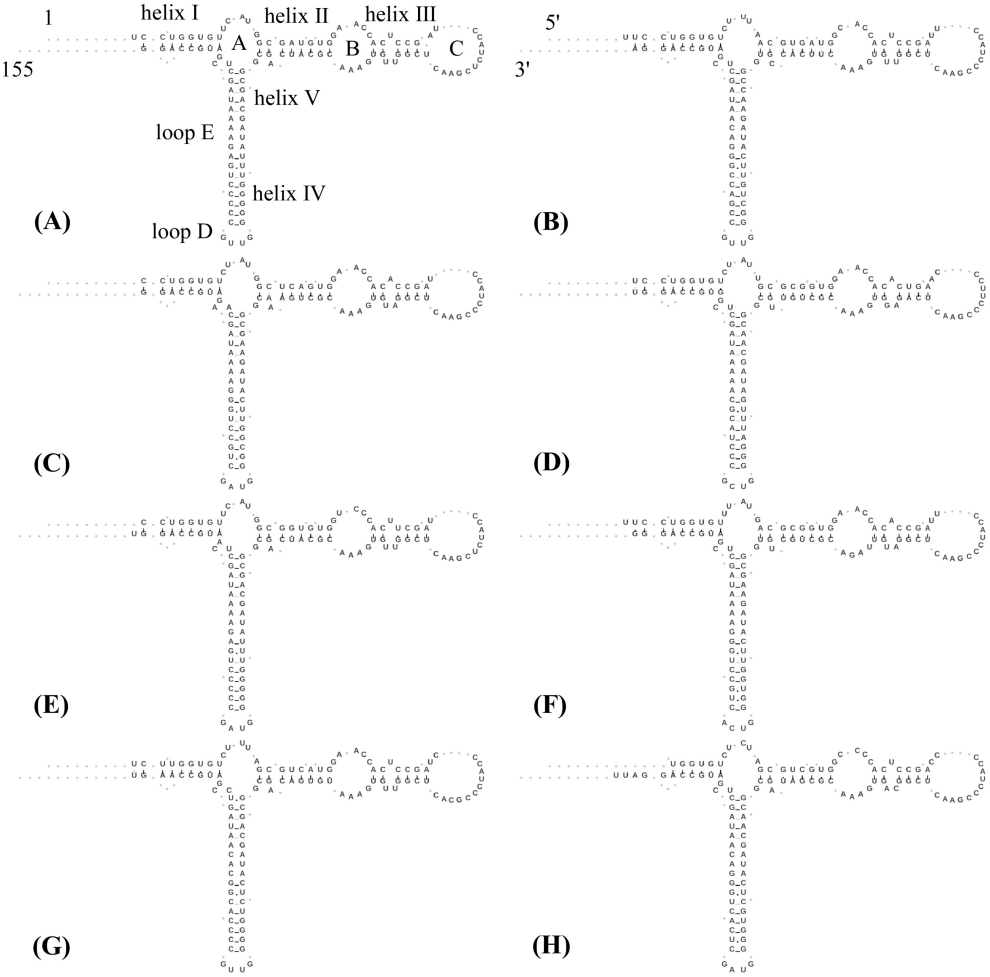


Fig. 3. Putative secondary structures of the 5S rRNAs. (A) *Prochlorococcus marinus*, (B) *Microcystis aeruginosa*, (C) *Arthrospira platensis*, (D) *Nostoc punctiforme*, (E) *Synechococcus sp.* WH 8102, (F) *Oscillatoria formosa* PCC 6407, (G) *Synechocystis sp.* PCC 6803 and (H) *Cyanothece sp.* PCC 8801

Conclusion. Our results demonstrated that the 5S rRNA and its secondary structure are suitable additional molecular markers for solving the taxonomic position and phylogenetic analyses of cyanobacterial strains.

REFERENCES

[1] Rippka R., J. Deruelles, J. B. Waterbury, M. Herdman, R. Y. Stanier (1979) Generic assignments, strain histories and properties of pure cultures of cyanobacteria, J. Gen. Microbiol., **111**, 1–61.

[2] Anagnostidis K., J. Komárek (1985) Modern approach to the classification system of the cyanophytes 1: Introduction, Algol. Stud., **38/39**, 291–302.
[3] Kondratieva E. N., N. Pfennig, H. G. Truper (1992) The phototrophic prokaryotes. In: The Prokaryotes (eds A. Balows, H. G. Trüper, M. Dworkin, W. Harder, K. H. Schleifer), New York, Springer, 312–330.
[4] Castenholz R. W. (2001) Phylum BX. Cyanobacteria: oxygenic photosynthetic bacteria. In: Bergey's Manual of Systematic Bacteriology, 2nd ed. (eds D. R. Boone, R. W. Castenholz, G. M. Garrity), New York, Springer, **1**, 474–487.
[5] Komárek J., J. Kaštovský, J. Mareš, J. R. Johansen (2014) Taxonomic classification of cyanoprokaryotes (cyanobacterial genera) 2014, using a polyphasic approach, Preslia, **86**, 295–335.
[6] Komárek J. (2006) Cyanobacterial taxonomy: current problems and prospects for the integration of traditional and molecular approaches, Algae, **21**(4), 349–375.
[7] Wilmotte A., M. Herdman (2001) Phylogenetic relationships among the cyanobacteria based on 16S rRNA sequences. In: Bergey's manual of systematic bacteriology, 2nd ed. (eds D. R. Boone, R. W. Castenholz, G. M. Garrity), Springer, New York, **1**, 487–493.
[8] Neilan B. A., D. Jacobs, A. E. Goodman, P. T. Cox, P. Hawkins (1995) Genetic diversity and phylogeny of toxic cyanobacteria determined by DNA polymorphism within the phycocyanin locus, Appl. Environ. Microbiol., **61**, 3875–3583.
[9] Boyer S. L., V. R. Flechtner, J. R. Johansen (2001) Is the 16S-23S rRNA internal transcribed spacer region a good tool for use in molecular systematics and population genetics? A case study in cyanobacteria, Mol. Biol. Evol., **18**, 1057–1069.
[10] Seo P.-S., A. Yokota (2003) The phylogenetic relationships of cyanobacteria inferred from 16S rRNA, gyrB, rpoC1 and rpoD1 gene sequences, J. Gen. Appl. Microbiol., **49**, 191–203.
[11] Teneva I., B. Dzhambazov, R. Mladenov, K. Schirmer (2005) Molecular and phylogenetic characterization of Phormidium species (Cyanoprokaryota) using the cpcB-IGS-cpcA locus, J. Phycology, **41**(1), 188–194.
[12] Stoyanov P., D. Moten, R. Mladenov, B. Dzhambazov, I. Teneva (2014) Phylogenetic relationships of some filamentous cyanoprokaryotic species, Evol. Bioinform. Online, **10**, 39–49.
[13] Zhaxybayeva O., J. P. Gogarten, R. L. Charlebois, W. F. Doolittle, R. T. Papke (2006) Phylogenetic analyses of cyanobacterial genomes: quantification of horizontal gene transfer events, Genome Res., **16**, 1099–1108.
[14] Shi T., P. G. Falkowski (2008) Genome evolution in cyanobacteria: the stable core and the variable shell, Proc. Natl. Acad. Sci., **105**, 2510–2515.
[15] Gupta R. S. (2009) Protein signatures (molecular synapomorphies) that are distinctive characteristics of the major cyanobacterial clades, Int. J. Syst. Evol. Microbiol., **59**, 2510–2526.
[16] Gupta R. S., D. W. Mathews (2010) Signature proteins for the major clades of Cyanobacteria, BMC Evol. Biol., **10**, 24.
[17] Shih P. M., D. Wu, A. Latifi, S. D. Axen, D. P. Fewer et al. (2013) Improving the coverage of the cyanobacterial phylum using diversity-driven genome sequencing, Proc. Natl. Acad. Sci. U.S.A., **110**, 1053–1058.
[18] Barciszewska M. Z., M. Szymanski, V. A. Erdmann, J. Barciszewski

(2001) Structure and functions of 5S rRNA, Acta Biochimica Polonica, **48**(1), 191–198.

[19] Kumar S., G. Stecher, K. Tamura (2016) MEGA7: Molecular Evolutionary Genetics Analysis v. 7.0 for bigger datasets, Mol. Biol. Evol., **33**, 1870–1874.

[20] Szymanski M., A. Zielezinski, W. M. Karlowski (2016) 5SRNAdb: an information resource for 5S ribosomal RNAs, Nucleic Acids Res., **44** (D1): D180-3, doi: 10.1093/nar/gkv1081.

[21] Kämpfer P., S. Glaeser (2012) Prokaryotic taxonomy in the sequencing era – the polyphasic approach revisited, Env. Microbiology, **14**(2), 291–317.

Department of Developmental Biology
**Department of Botany*
Faculty of Biology
Plovdiv University "Paisii Hilendarski"
24, Tsar Assen St
4000 Plovdiv, Bulgaria
e-mails: moten@uni-plovdiv.bg
diyana_basheva@abv.bg
rummlad@uni-plovdiv.bg
balik@uni-plovdiv.bg
teneva@uni-plovdiv.bg