ACTA BIOLOGICA SLOVENICA
LJUBLJANA 2022 Vol. 65, Št. 2: 104–115

Phylogenetic study of *Aliinostoc* species (Cyanobacteria) using *pc-igs*, *nifH* and *mcy* as markers for investigation of horizontal gene transfer

Filogenetska študija vrst *Aliinostoc* (Cyanobacteria) z uporabo označevalcev *pc-igs*, *nifH* in *mcy* za ugotavljanje horizontalnega genskega prenosa

Bahareh Nowruzi

Department of Biotechnology, Science and Research Branch, Islamic Azad University, Tehran, Iran.

Correspondence: bahare77biol@gmail.com

Abstract: Selection of genes that have not been horizontally transferred for prokaryote phylogenetic studies is regarded as a challenging task. Internal transcribed spacer of ribosomal genes (16S–23S ITS), microcystin synthetase genes (*mcy*), nitrogenase (*nifH*) and phycocyanin intergenic spacer (*PC-IGS*) are among the most used markers in cyanobacteria. The region of the ribosomal genes has been considered stable, whereas the *nifH*, *mcyG* and *PC-IGS* may have undergone horizontal transfer. To investigate the occurrence of horizontal transfer of *nifH*, *mcyG* and *PC-IGS*, phylogenetic trees of *Aliinostoc* strains Ay1375 and Me1355 were generated and compared. Phylogenetic trees based on the markers were mostly congruent for *PC-IGS*, indicating a common evolutionary history among ribosomal and phycocyanin genes with no evidence for horizontal transfer of *PC-IGS*. Phylogenetic trees constructed from the *nifH* and 16S rRNA genes were incongruent. Our results suggest that *nifH* has been transferred from one cyanobacterium to another. Moreover, the low non-synonymous/synonymous mutation ratio (Ka/Ks) was consistent with an ancient origin of the *mcyG*.

Keywords: 16S–23S ITS, cyanobacteria, horizontal gene transfer, molecular phylogeny, phycocyanin, ribosomal genes

Izvleček: Za filogenetske študije prokariontov velja, da je izbira genov, ki niso bili horizontalno preneseni, zahtevna naloga. Notranji prepisani vmesnik ribosomskih genov (16S–23S ITS), geni mikrocistin sintetaze (*mcy*), nitrogenaze (*nifH*) in fikocianinski medgenski vmesnik (*PC-IGS*) so med najpogosteje uporabljenimi označevalci pri cianobakterijah. Območje ribosomskih genov velja za stabilno, medtem ko so zaporedja *nifH*, *mcyG* in PC-IGS lahko bila prenesena s horizontalnim genskim prenosom. Da bi raziskali pojav horizontalnega prenosa *nifH*, *mcyG* in PC-IGS, smo ustvarili filogenetska drevesa sevov Ay1375 in Me1355 vrste *Aliinostoc* ter jih med seboj primerjali. Filogenetska drevesa na podlagi označevalcev so bila večinoma skladna za *PC-IGS* in niso razkrila morebitnih horizontalnih genskih prenosov, kar kaže na skupno evolucijsko zgodovino med ribosomskimi in fikocianinskimi geni. Primerjava filogenetskih dreves, pridobljenih na podlagi gena *nifH* s filogenetskimi drevesi, pridobljenimi na podlagi gena za 16S rRNA, je razkrila neskladja. Naši rezultati tako nakazujejo, da je bil *nifH* prenesen iz ene cianobakterije v drugo s horizontalnim genskim prenosom. Poleg tega

se nizko razmerje med nesinonimnimi/sinonimnimi mutacijami (Ka/Ks), ki smo ga razkrili v študiji, sklada s starodavnim izvorom gena *mcyG*.

Ključne besede: 16S–23S ITS, cianobakterije, horizontalni genski prenos, molekularna filogenija, fikocianin, ribosomski geni

Introduction

The morphological characteristics of cyanobacteria do not always correspond to their taxonomic diversity (Komárek et al. 2016) and therefore the use of molecular markers for phylogenetic studies have become essential (Han et al. 2009).

Aliinostoc species is a cosmopolitan, nitrogen (N_2)-fixing cyanobacterial species found in temperate to tropical freshwater or terrestrial habitats. The widespread proliferation of *Aliinostoc* species in paddy fields has increased the nitrogen in soils. Molecular approaches are particularly useful in the detection and identification of specific strains, especially those that are morphologically identical at the species level. Genetic identification can also be used to characterize the degree of genetic similarity among populations (Kabirnataj et al. 2020; Nowruzi et al., 2021; Nowruzi and Shalygin 2021).

One of the genes utilized for genetic differences between *Aliinostoc* cultures was *nifH*, a highly conserved gene that encodes dinitrogenase reductase, a protein subunit in the nitrogenase complex involved in N_2 fixation. Common to all N_2 fixers, the 324-bp *nifH* fragment is useful in characterizing diazotrophic communities and for differentiating cyanobacterial genera (Foster and Zehr, 2006). The other genetic locus used was *cpcBA*-IGS, which includes the highly variable intergenic spacer (IGS) region between two phycobilisome subunits (*cpcB* and *cpcA*) within the phycocyanin operon (Dyble et al., 2002; Brient et al., 2008). Both *cpcA*-IGS (Bastien et al., 2011) and *nifH* appear to be more useful in discriminating between strains than the commonly employed 16S rRNA gene, which exhibits low intrageneric variability in many cyanobacteria (Teneva et al., 2012). Moreover, microcystins, cyclic heptapeptide hepatotoxins, are by far the most prevalent of the cyanobacterial toxins and are produced by microcystin synthetase gene cluster (Jungblut et al. 2006; Nowruzi et al., 2022).

One of the greatest challenges in the selection of markers for phylogenetic studies in cyanobacteria is targeting markers that have not undergone horizontal gene transfer (HGT) (Yerrapragada and Siefert, 2009; Piccin-Santos et al., 2014). HGT and orthologous gene substitutions are relatively common among cyanobacteria and have been important processes in the evolution of this group (Piccin-Santos et al., 2014). However, HGT events in cyanobacteria may still be underestimated, and genes with several functions could have been subjected to this process (Zhaxybayeva et al. 2006). There is no reported evidence that the operons of ribosomal genes have undergone HGT among cyanobacteria. However, the variability observed among the multiple copies of the ribosomal operon found within a single individual can hinder their use in phylogenetic studies (Iteman et al. 2002).

The construction and comparison of phylogenetic trees are perhaps the best ways to assess the contribution of HGT to the evolutionary history of a gene family (Koonin et al., 2002). Incongruence is taken to indicate a role for HGT, whereas congruence is consistent with descent through common ancestry. Therefore, to resolve the relationship between microcystin synthetase genes, *PC-IGS*, *nifH*, 16S rRNA and the role of HGT in the evolutionary history, we undertook a molecular phylogenetic study. We analyzed and tested for congruence two data sets comprised of genes involved in primary metabolism and genes involved directly in the synthesis of microcystins and nodularins.

Our goal, using strains of *Aliinostoc* species as models, was to evaluate the possible occurrence of HGT by comparing phylogenetic trees built with *mcy*, *PC-IGS*, *nifH* and 16S rRNA. This is the first study to compare the different molecular markers in characterizing two *Aliinostoc* isolates originating from paddy fields of Iran.

Abbreviations: 16S–23S ITS, internal transcribed spacer of ribosomal genes marker; HGT, horizontal gene transfer; PC-IGS, phycocyanin intergenic spacer marker

Material and methods

Strains and cultivation conditions

The clonal and axenic strains (strain designations Ay1375 and Me1355) of *Aliinostoc* belonged to the Cyanobacteria Culture Collection (CCC) and ALBORZ herbarium. Strains were maintained in climate chambers with controlled conditions of continues light and temperature (25 ± 5°C) in BG-11 cultivation medium (Rippka et al. 1979), of pH value 7.4.

Molecular and sequence analysis

Genomic DNA was isolated from 16-18 days old log phase cultures using the Himedia Ultrasensitive Spin Purification Kit (MB505) following the instructions of the manufacturer, except the increase of incubation time for the lysis solutions AL and C1, which were set to 60 and 20 min, respectively. DNA fragments within the following genes were amplified using the oligonucleotide primers and PCR programs listed in Table 1: 16S rRNA gene, ITS, *nifH*, *PC-IGS, mcyG and mcyD.* PCR reactions were performed using a thermal cycler 5.9 and the following procedure: 25 µl aliquots containing 10-20 ng DNA template, 0.5 µM of each primer, 1.5 mM $MgCl_2$, 200 µM dNTPs and 1U/µl Taq DNA polymerase (Robertson et al., 2001; Dyble et al., 2002; Nowruzi and Lorenzi, 2021). PCR products were analyzed by electrophoresis on 1% agarose gels (SeaPlaque® GTG®, Cambrex Corporation), using standard protocols. The products were purified directly using the Geneclean® Turbo kit (Qbiogene, MP Biomedicals) and sequenced using the BigDye® Terminator v3.1 cycle sequencing kit (Applied Biosystems, Life Technologies).

The partial sequences were compared with the ones available in the NCBI database (March, 2022) using BLASTn. The BLAST X tool (blast.ncbi.nlm.nih.gov/Blast.cgi) was used for *cpc*A-IGS, *nifH, mcy D* and *mcyE genes.* The sequences were annotated with the NCBI ORF Finder and the ExPASY (https://www.ncbi.nlm.nih.gov/orffinder/) proteomics tools.

Table 1. Target genes, oligonucleotide primers and PCR programs used in this study.
Tabela 1: Tarčni geni, začetni oligonukleotidi in programi PCR, uporabljeni v raziskavi.

Target gene/ sequence	Primer designation (sequence 5´→3´)	PCR program (reference in superscript)
16S rRNA	PA (5'-AGAGTTTGATCCTGGCTCAG-3') B23S (5'-CTTCGCCTCTGTGTGCCTAGGT-3')	[1]94˚C, 3 min [1]30 × (94˚C, 30 s; 55˚C, 40 s; 72˚C, 1.30 min) [2]72˚C, 3 min [2]4˚C, ∞
16S-23S rRNA ITS	ITS-F (5'-TGTACACACCGCCCGTC-3') ITS-R (5'-CTCTGTGTGCCTAGGTATCC-3')	
***cpc*A-IGS**	Cpc F (5'-GGCTGCTTGTTTACGCGACA-3') Cpc R (5'-CCAGTACCACCAGCAACTAA-3')	[3]94˚C, 5 min [3]30 × (92˚C, 1 min; 55˚C, 1 min; 72˚C, 2 min) [3]72˚C, 6 min [3]4˚C, ∞
psbA	PSBA86F (5'-TTTATGTGGGTTGGTTCGG-3') PSBA980R4 (5'-TGAGCATTACGCTCGTGC-3')	[4]94˚C, 5 min [4]35 × (94˚C, 60 s; 56˚C, 60 s; 72˚C, 60 s) [5]72˚C, 10 min [5]4˚C, ∞
***nif*H**	*nif*H F (5'-CGTAGGTTGCGACCCTAAGGCTGA-3') *nif*H R (5'-GCATACATCGCCATCATTTCACC-3')	
mcyG	*mcyG* F (5'-GAAATTGGTGCGGGAACTGGAG-3') *mcyG* R (5'-TTTGAGCAACAATGATACTTTGCTG-3')	[6]95˚C, 5 min [6]34 × (95˚C, 30 s; 53˚C, 30 s; 72˚C, 60 s) [7]72˚C, 5 min [7]4˚C, ∞
***mcy*D**	*mcy*D F (5'-GCTCAAGAAAAATTACATCAAG-3') *mcy*D R (5'-TTAAAGGAGAATGAAAAGCATGAGA-3')	

References: [1]Taton et al. 2003; [2]Iteman et al. 2000, [3]Neilan et al. 1997, [4] Junier et al. 2007, [5]Gaby and Buckley 2012, [6] Fewer et al. 2007, [7] Rantala et al. 2004

Nucleotide sequence accession numbers

Sequence data were deposited in the DNA Data Bank of Japan (DDBJ) under the accession numbers showed in Table 2.

Table 2: Accession numbers of sequence data deposited in the DNA Data Bank of Japan. *mcyG* was not found in *Aliinostoc* sp. Ay1375.
Tabela 2: Oznake v japonski podatkovni bazi DNA deponiranih nukleotidnih zaporedij *Aliinostoc* sp. *mcyG* pri *Aliinostoc* sp. Ay1375 ni bil določen.

Nucleotide ID	Target gene	Strain	Number of nucleotides	Number of amino acids	Tree model
ON751925	*16S rRNA*	Ay1375	1442	-	TVM+F+I+G4
ON751926		Me1355	1442		
ON755128	*nifH*	Ay1375	286	80	TIM2e+I+G4+F
ON755129		Me1355	282	79	
ON755126	*cpcA*-IGS	Ay1375	589	98	LG+G4
ON755127		Me1355	612	98	
OM801556	*mcyG*	Me1355	494	164	TPM3U+G4+F

Phylogenetic analysis

The 16S rRNA, ITS, *cpc*A-IGS, *nifH, mcyG* and *mcyD* genes sequences obtained in this study, as well as the best hit sequences (> 94% identity) retrieved from GenBank, were first aligned using MUSCLE (Edgar 2004), and then maximum likelihood phylogenetic trees were inferred in IQ-Tree (multicore v1.5.5) (Nguyen et al. 2015). Different models were used as suggested (BIC criterion) after employing model test implemented in IQ-tree (Table 2). Tree robustness was estimated with bootstrap percentages using 100 standard bootstrap and 10,000 ultrafast bootstrap to evaluate branch supports (Guajardo-Leiva et al. 2018).

16S-23S rRNA ITS region secondary structure analysis

The sequences corresponding to the D1-D1' helix, D2, D3, Box-B and Box-A regions of the 16S-23S ITS of the studied strains were characterized according to the Johansen et al. (2011), and $trRNA^{Ile}$ and $trRNA^{Ala}$ were determined according to the tRNAscan-SE 2.0 (Chan et al., 2021). Comparison of the ITS secondary structures of studied strains and the reference strains were generated using the M-fold web server (version 2.3) (Zuker 2003) under ideal conditions of untangled loop fix and the temperature set to default (37 °C).

Sequence divergence

We calculated the number of non-synonymous substitutions per non-synonymous site (Ka) and the number of synonymous substitutions per synonymous site (Ks) by using MEGA X (Nowruzi and Blanco, 2019). A Ka/Ks ratio >1 indicates positive selection for advantageous mutations, whereas a Ka/Ks ratio <1 indicates purifying selection to prevent the spread of detrimental mutations (Leikoski et al., 2009).

Results

Phylogenetic analyses

Phylogenetic trees based on different gen markers are shown in Figs. 1 to 3. The *nifH* gene fragment and the fragment of the phycocyanin operon (*cpcA*-IGS) were amplified from both studied strains, however *mcyG* was only detected in *Aliinostoc* Me1355 strain.

The *Aliinostoc* phylogenetic trees based on the markers *PC-IGS* and 16S–23S ITS (Fig. 1) showed similar topologies. From the phylogenetic analysis based on 16 rRNA gene sequences, it is possible to observe that the studied strain is within a cluster composed by other *Nostoc* strains and its closest one is *Nostoc_elgonense*_TAU-MAC_0299 (MN062664). However, the phylogenetic trees obtained using *nifH* and 16S–23S ITS (Fig. 2) have differences in the branch positions of some strains. In the phylogeny based on the gene 16S–23S ITS, the studied strains were placed with *Nostoc_calcicola*_Ind32 (N216874) in the same cluster, However, when we look into the phylogeny based on *nifH* gene, the studied strains fall into separate clades and its closest one is *Nostoc*_sp. NQAIF320 (KJ636979), indicating that this gene probably could be the best marker for a high resolution at species level.

The highest *mcyG* sequence similarity was found to be 100% identical with *Nostoc* sp. CENA88 (Q259210) (Fig. 3). Moreover, The *Aliinostoc* phylogenetic trees based on the markers *mcyG* and 16S–23S ITS (Fig. 3) showed similar topologies. We have also compared the 16S rRNA *p*-distances of our strains with related genera, namely *Aliinostoc_morphoplasticum* NOS (KY403996_1), *Aliinostoc* sp. SA46 (MK503795), *Aliinostoc* sp. SA9 (MK503790) and *Aliinostoc magnakinetifex* SA18 (MK503791). Results showed that *Aliinostoc* sp. strain Ay1375 shared a 16S rRNA sequence similarity of 97.22% with *Aliinostoc morphoplasticum* NOS (KY403996_1), 96.52% with *Aliinostoc* sp. SA46 (MK503795), 96.80% with *Aliinostoc* sp. SA9 (MK503790) and 93.73% with *Aliinostoc magnakinetifex* SA18 (MK503791), while *Aliinostoc* sp. strain Me1355 shared a 16S rRNA sequence similarity of 97.22% with *Aliinostoc morphoplasticum* NOS (KY403996_1), 96.38% with *Aliinostoc* sp. SA46 (MK503795), 96.80% with *Aliinostoc* sp. SA9 (MK503790) and 93.59% with *Aliinostoc magnakinetifex* SA18 (MK503791) (Tab. 3).

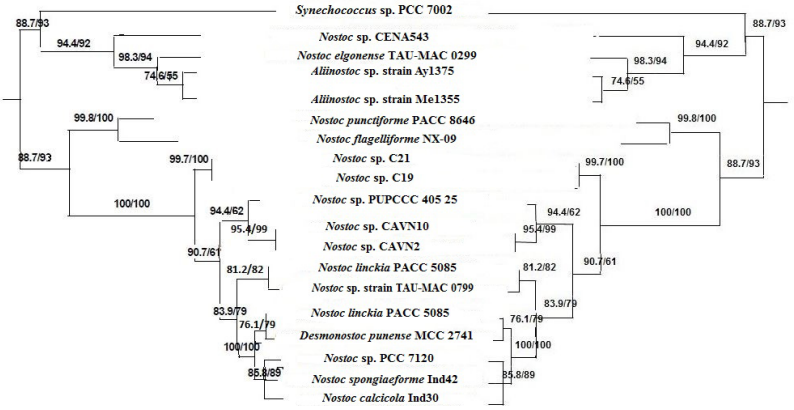

Figure 1: Congruence between phylogenies inferred from the 16S rRNA and *cpc*A-IGS sequences. A maximum-likelihood tree based on the 16SrRNA data set (left). A maximum-likelihood tree based on the *cpc*A-IGS data set (right). Numbers near nodes indicate standard bootstrap support (%) / ultrafast bootstrap support (%) for ML analyses.

Slika 1: Skladnost med nizoma podatkov nukleotidnih zaporedij 16S rRNA in *cpc*A-IGS. Drevo, ocenjeno po metodi največjega verjetja na osnovi 16S rRNA (levo). Drevo, ocenjeno po metodi največjega verjetja na osnovi *cpc*A-IGS (desno). Številke ob razvejitvah prikazujejo podporo izračunano z običajno metodo samovzorčenja (%) / ultrahitro metodo samovzorčenja (%) za analize največjega verjetja.

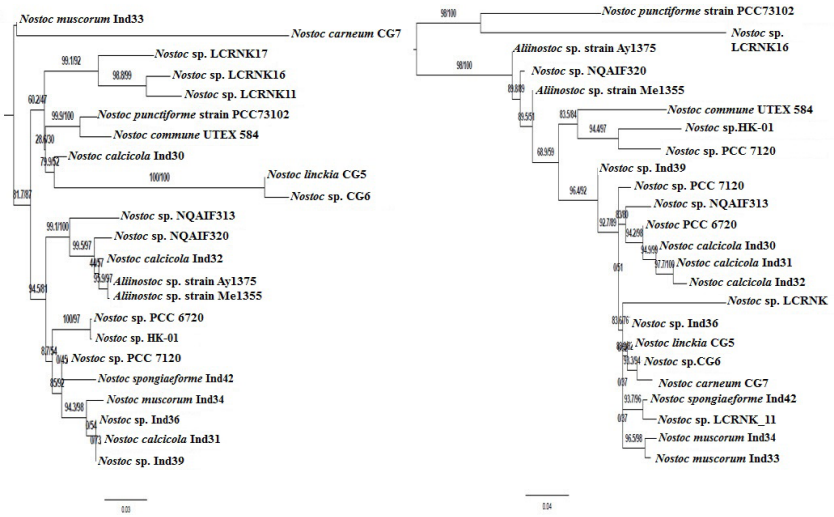

Fig. 2. Congruence between phylogenies inferred from the 16S rRNA and *nifH* sequences. A maximum-likelihood tree based on the 16SrRNA data set (left). A maximum-likelihood tree based on the *nifH* data set (right). Numbers near nodes indicate standard bootstrap support (%) / ultrafast bootstrap support (%) for ML analyses.

Slika 2: Skladnost med nizoma podatkov nukleotidnih zaporedij 16S rRNA in *nifH*. Drevo, ocenjeno po metodi največjega verjetja na osnovi 16S rRNA (levo). Drevo, ocenjeno po metodi največjega verjetja na osnovi *nifH* (desno). Številke ob razvejitvah prikazujejo podporo izračunano z običajno metodo samovzorčenja (%) / ultrahitro metodo samovzorčenja (%) za analize največjega verjetja.

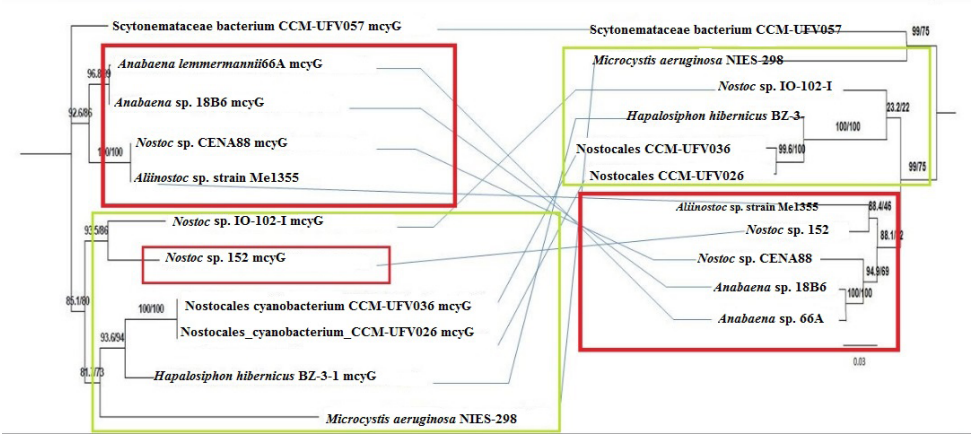

Fig. 3. Congruence between phylogenies inferred from the 16S rRNA and *mcyG* sequences. A maximum-likelihood tree based on the *mcyG* data set (left). A maximum-likelihood tree based on the 16SrRNA data set (right). Numbers near nodes indicate standard bootstrap support (%) / ultrafast bootstrap support (%) for ML analyses.

Slika 3: Skladnost med nizoma podatkov nukleotidnih zaporedij 16S rRNA in *mcyG*. Drevo, ocenjeno po metodi največjega verjetja na osnovi 16S rRNA (levo). Drevo, ocenjeno po metodi največjega verjetja na osnovi *mcyG* (desno). Številke ob razvejitvah prikazujejo podporo izračunano z običajno metodo samovzorčenja (%) / ultrahitro metodo samovzorčenja (%) za analize največjega verjetja.

Table 3. 16S rRNA gene sequence similarity matrix of studied strains and related taxa.
Tabela 3: Matrika podobnosti za nukleotidno zaporedje gena 16S rRNA pri preučevanih sevih in sorodnih taksonih.

Strain	*Aliinostoc* sp. strain Ay1375	*Aliinostoc* sp. strain Me1355	KY403996_1_ *Aliinostoc_ morphoplasticum_* NOS	MK503795_1_ *Aliinostoc_* sp_SA46	MK503790_1_ *Aliinostoc_* sp_SA9
***Aliinostoc* sp. strain Ay1375**					
***Aliinostoc* sp. strain Me1355**	0				
***Aliinostoc_morphoplasticum*_NOS**	97.22	97.22			
***Aliinostoc*_sp_SA46**	96.52	96.38	95.80		
***Aliinostoc*_sp_SA9**	96.80	96.80	95.52	99.33	
***Aliinostoc_magnakinetifex*_SA18**	93.73	93.59	94.16	95.79	95.50

16S-23S rRNA ITS secondary structure

Four reference sequences were used to search for ITS secondary structure. According to Johansen et al. (2011), nine different areas (D1-D1' helix, D2, D3, trRNA^{Ile}, trRNA^{Ala}, Box-B, Box-A and D4) were found in the ITS secondary structure of studied strain. The D1-D1' and Box-B regions of all studied strains were revealed to be very different in terms of length and shape (Fig. 4, Tab. 4).

The D1-D1' region included a terminal bilateral bulge (A), bilateral bulge (B), unilateral bulge (C), and basal clamp (D) (Fig. 4). The lengths of D1-D1' helix varied from 93 nt (*Aliinostoc* sp. strain Ay1375, *Aliinostoc* sp. strain Me1355, KY403996.1 *Aliinostoc morphoplasticum* NOS) to 60 nt (*Aliinostoc magnakinetifex* SA18) (Tab. 4). The basal stem revealed to be the same for all studied strains (5'- GACCUA- UAGGUC - 3') (Fig. 4).

Box-B was nominated by a terminal bilateral bulge (A) and bilateral bulge (B). Box-B helix was not found for *Aliinostoc magnakinetifex* SA18. As to the Box-B + spacer, lengths varied from 39 nt (*Aliinostoc morphoplasticum* NOS) to 55 nt (*Aliinostoc* sp. SA46), with studied strains showing a length of 44 nt (Fig. 4) (Tab. 5).

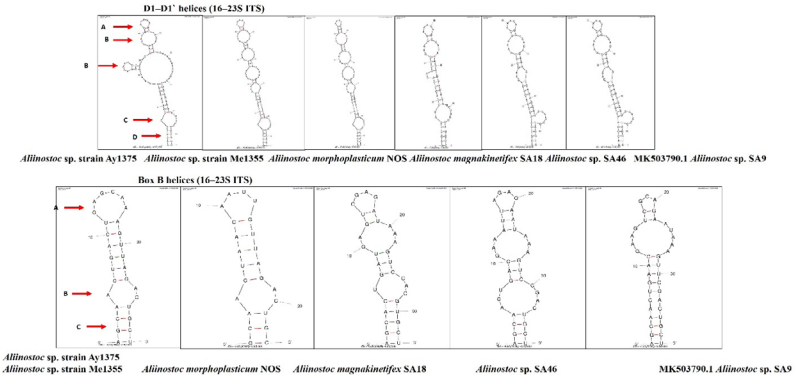

Figure 4. Comparison of secondary structures of D1–D1′ helices (upper row) and Box-B helices and V3 helices (lower row), both from 16S–23S intergenic spacers between studied strains with reference strains. Marks: A - Terminal bilateral bulge, B - bilateral bulge, C - Unilateral bulge, D - Basal clamp, arrows - bulges and basal clamp.
Slika 4: Primerjava sekundarnih struktur D1-D1' vijačnic (zgornja vrstica) ter Box-B vijačnic in V3 vijačnic (spodnja vrstica) iz 16S–23S medgenskih vmesnikov med preučevanimi in referenčnimi sevi. Oznake: A - Terminalna bilateralna izboklina, B - Bilateralna izboklina, C - Unilateralna izboklina, D – Bazalna spona, puščice – izbokline in bazalna spona.

Table 4: Nucleotide lengths of the 16S–23S ITS regions of the studied strains.
Tabela 4: Dolžine nukleotidov za območja 16S–23S ITS pri preučevanih sevih.

Strain	D1-D1' helix	spacer+D2+spacer	D3	D3 + spacer	trRNA^{Ile} gene	spacer+V2+spacer	TrRNA^{Ala} gene	BoxB+spacer	Box A	D 4
***Aliinostoc* sp. strain Ay1375**	93	39	3	30	-	-	-	44	11	9
***Aliinostoc* sp. strain Me1355**	93	39	3	30	-	-	-	44	11	9
***Aliinostoc morphoplasticum* NOS**	93	38	3	22	-	-	-	39	11	9
***Aliinostoc magnakinetifex* SA18**	60	34	3	37	-	-	-	-	-	9
***Aliinostoc* sp. SA46**	66	38	3	43	-	-	-	55	10	10
***Aliinostoc* sp. SA9**	66	40	3	46	-	-	-	54	11	9

Table 5: Comparison of secondary structure of 16S-23S rRNA (D1-D1'helix and Box-B helix) between the studied strains with reference strains.
Tabela 5: Primerjava sekundarne zgradbe 16S-23S rRNA (vijačnica D1-D1' in Box-B) med preučevanimi in referenčnimi sevi.

Strain	D1-D1'helix				Box-B		
	Terminal bilateral bulge (A)	Bilateral bulge (B)	Unilateral bulge (C)	Basal clamp (D)	Terminal bilateral bulge (A)	Bilateral bulge (B)	Basal clamp (C)
	Number of nucleotides	Number of loops	Number of loops	Number of nucleotides	Number of nucleotides	Number of loops	Number of nucleotides
***Aliinostoc* sp. strain Ay1375**	7	3	1	12	6	1	8
***Aliinostoc* sp. strain Me1355**	7	4	1	12	6	1	8
***Aliinostoc morphoplasticum* NOS**	6	4	1	12	8	1	6
***Aliinostoc magnakinetifex* SA18**	7	2	1	12	6	2	10
***Aliinostoc* sp. SA46**	7	1	2	12	6	2	8
***Aliinostoc* sp. SA9**	7	1	2	12	5	1	22

Sequence divergences

Sequence divergences in the *mcyG* gene data set were much higher than expected in an evolutionary scenario, favoring recent horizontal gene transfer as a mechanism to explain the sporadic distribution of microcystin producers among cyanobacteria. To determine whether the *mcyG* gene is under positive

or negative selection pressure, we compared the number of nonsynonymous substitutions per non-synonymous site (Ka) to the number of synonymous substitutions per synonymous site (Ks). The Ka/Ks ratio was well below 1 in pairwise comparisons from representative strains of each genus. A low Ka/Ks ratio is indicative of purifying selection in which deleterious mutations affecting the protein sequence are selected against and is consistent with an ancient origin of the *mcyG* gene.

Discussion

HGT is relatively common among cyanobacteria, but it does not affect all genes in the same way. For some genomes, gene clusters have a lower probability of being transferred (Rantala et al., 2004).

The phylograms based on *PC-IGS* and *mcyG* were mostly congruent and no clear HGT signal was found for these genes, indicating a common evolutionary pathway for the phycocyanin, *mcyG* and ribosomal genes. This result is consistent with those of Sanchis et al. (2005) and Dadheech et al. (2010), who found that *PC-IGS* and 16S–23S ITS regions of *Microcystis* and *Arthrospira* strains also showed a high similarity between marker topologies. Phylogenetic analysis of this region was largely consistent with that obtained from 16S rDNA sequence analysis and revealed a relationship between the 16S rDNA sequence and the phycobilin content of cells.

However, phylogenetic trees constructed from the *nifH* and 16S rRNA genes were incongruent. Our results suggest that the *nifH* gene encoding the dinitrogenase reductase has been transferred from one cyanobacterium to another. However, the phylogenetic incongruence detected is likely to be a result of ancient horizontal transfers of the *nifH* biosynthetic genes since the sequence divergence of the dinitrogenase reductase genes was high. The main point of discordance between in the *nifH* phylogenetic tree was the location of two studied strains, in the phylogeny based on the gene 16S–23S ITS, they were placed in the same cluster, however into the phylogeny based on *nifH* gene, the studied strain falls into separate clades. In addition, it is noteworthy that in the 16S–23S ITS proposed phylogeny, there is a high Bayesian posterior probability to support its location.

Morphological studies showed that both studied strains were morphologically similar to each other, but that two of them formed an isolated clade in the *nifH* phylogram, indicating that despite the morphological similarity, they represent genetically divergent strains. Thus, the hypothesis that the divergence of the strains observed in the *nifH* tree could have been due to HGT was confirmed.

Moreover, our analyses do not corroborate the presence of HGT in *PC-IGS* and *mcyG*, but this event cannot be neglected as a hypothesis for explaining divergences in phylogenies. A study on the genome of Synechococcus spp. indicated that genes encoding phycocyanin may have evolved independently from genes of the core genome such as the *allo-PC* gene or the ribosomal regions (Six et al. 2007).

The search for more stable markers, not biased by HGT, has become essential for understanding the phylogeny and taxonomy of cyanobacteria (Gribaldo and Brochier 2009). The results presented herein strongly support *nifH* as a marker of choice for cyanobacterial phylogenetic studies and emphasize the importance of using multiple molecular markers to prevent erroneous conclusions based on HGT.

Summary

Horizontal gene transfer (HGT), potentially followed by recombination with or replacement of resident homologues, represents an important factor in the phylogeny of prokaryotic organisms such as cyanobacteria, and shapes their evolutionar history. Nowadays, HGT seems to be a major factor in species delimitation in cyanobacteria and plays a key selection pressure leading to cyanobacterial diversification. In this study, *PC-IGS*, *nifH*, *mcyD, mycG* and the ribosomal gene spacer 16S–23S ITS as molecular markers were compared to investigate the occurrence of horizontal transfer. The phylograms based on *PC-IGS* and *mcyG* were mostly congruent and no clear HGT signal was found for these genes. However, phylogenetic trees constructed from the *nifH* and 16S rRNA genes were incongruent. The exploration for more

steady markers, not biased by HGT, has become important for detection of the phylogeny and taxonomy of cyanobacteria.

Povzetek

Horizontalni genski prenos (HGT), ki mu lahko sledi rekombinacija ali zamenjava obstoječih homolognih zaporedij, predstavlja pomemben dejavnik v filogeniji prokariontskih organizmov, kot so cianobakterije, in oblikuje njihovo evolucijsko zgodovino. Danes se zdi, da je HGT glavni dejavnik pri razmejitvi vrst pri cianobakterijah in je ključni selekcijski pritisk, ki vodi v diverzifikacijo cianobakterij. V tej študiji smo primerjali nukleotidna zaporedja *PC-IGS*, *nifH*, *mcyD*, *mycG* in ribosomski medgenski vmesnik 16S–23S ITS kot molekularne označevalce, da bi raziskali pojav HGT. Filogenetska drevesa, ki temeljijo na nukleotidnih zaporedijh *PC-IGS* in *mcyG* so si med seboj bila večinoma skladna, tzato lahko za ta nukleotidna zaporedja predvidevamo, da se niso prenašala s HGT. Filogenetska drevesa, ki so bila narejena na podlagi nukleotidnih zaporedij *nifH* in genov za 16S rRNA, so bila med seboj neskladna, kar nakazuje na HGT. Raziskovanje bolj stabilnih označevalcev, na katere HGT ne vpliva, je postalo pomembno za odkrivanje filogenije in taksonomije cianobakterij.

References

Bastien, C., Cardin, R., Veilleux, É., Deblois, C., Warren, A., Laurion, I., 2011. Performance evaluation of phycocyanin probes for the monitoring of cyanobacteria. Journal of Environmental Monitoring. 3, 110-8.

Brient, L., Lengronne, M., Bertrand, E., Rolland, D., Sipel, A., Steinmann, D., Baudin, I., Legeas, M., Le Rouzic, B., Bormans, M., 2008. A phycocyanin probe as a tool for monitoring cyanobacteria in freshwater bodies. Journal of Environmental Monitoring. 10, 248-55.

Chan, P.P., Lin, B.Y., Mak, A.J., Lowe, T.M., 2021. tRNAscan-SE 2.0: improved detection and functional classification of transfer RNA genes. Nucleic Acids Research, 16, 9077-96.

Dadheech, P.K., Ballot, A., Casper, P., Kotut, K., Novelo, E., Lemma, B., Pröschold, T., Krienitz, L., 2010. Phylogenetic relationship and divergence among planktonic strains of *Arthrospira* (Oscillatoriales, Cyanobacteria) of African, Asian and American origin deduced by 16S–23S ITS and phycocyanin operon sequences. Phycologia, 4, 361-72.

Dyble, J., Paerl, H.W., Neilan, B.A., 2002. Genetic characterization of Cylindrospermopsis raciborskii (Cyanobacteria) isolates from diverse geographic origins based on nifH and cpcBA-IGS nucleotide sequence analysis. Applied and Environmental Microbiology, 5, 2567-71.

Edgar, R.C., 2004. MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Research, 32, 1792-7.

Fewer, D.P., Rouhiainen, L., Jokela, J., Wahlsten, M., Laakso, K., Wang, H., Sivonen, K., 2007. Recurrent adenylation domain replacement in the microcystin synthetase gene cluster. BMC Evolutionary Biology, 7, 183.

Foster, R.A., Zehr, J.P., 2006. Characterization of diatom–cyanobacteria symbioses on the basis of nifH, hetR and 16S rRNA sequences. Environmental Microbiology, 11, 1913-25.

Gaby, J.C., Buckley, D.H., 2012. A comprehensive evaluation of PCR primers to amplify the nifH gene of nitrogenase. PloS One, 7, 3-6.

Gribaldo, S., Brochier, C., 2009. Phylogeny of prokaryotes: does it exist and why should we care? Research in Microbiology, 7, 513-21.

Guajardo-Leiva, S., Pedrós-Alió, C., Salgado, O., Pinto, F., Díez, B., 2018. Active crossfire between Cyanobacteria and Cyanophages in phototrophic mat communities within hot springs. Frontiers in Microbiology, 3, 20-39.

Han, D., Fan, Y., Hu, Z., 2009. An evaluation of four phylogenetic markers in *Nostoc*: implications

for cyanobacterial phylogenetic studies at the intrageneric level. Current Microbiology, 58, 170-6.
Iteman, I., Rippka, R., de Marsac, N.T., Herdman, M., 2000. Comparison of conserved structural and regulatory domains within divergent 16S rRNA–23S rRNA spacer sequences of cyanobacteria. The GenBank accession numbers for the sequences reported in this paper are AF180968 and AF180969 for ITS-L and ITS-S, respectively. Microbiology, 146, 1275-86.
Iteman, I., Rippka, R., de Marsac, N.T., Herdman, M., 2002. rDNA analyses of planktonic heterocystous cyanobacteria, including members of the genera *Anabaenopsis* and *Cyanospira*. The GenBank accession numbers of the 16S rDNA gene sequences reported in this paper are AY038032–AY038037. Microbiology, 148, 481-96.
Johansen, J.R., Kovacik, L., Casamatta, D.A., Iková, K.F., Kastovský, J., 2011. Utility of 16S-23S ITS sequence and secondary structure for recognition of intrageneric and intergeneric limits within cyanobacterial taxa: *Leptolyngbya corticola* sp. nov.(*Pseudanabaenaceae*, Cyanobacteria). Nova Hedwigia, 92, 283.
Jungblut, A.D., Neilan, BA., 2006. Molecular identification and evolution of the cyclic peptide hepatotoxins, microcystin and nodularin, synthetase genes in three orders of cyanobacteria. Archives of Microbiology, 185, 107-14.
Junier, P., Witzel, K.P., Hadas, O., 2007. Genetic diversity of cyanobacterial communities in Lake Kinneret (Israel) using 16S rRNA gene, psbA and ntcA sequence analyses. Aquatic Microbial Ecology, 29, 233-41.
Kabirnataj, S., Nematzadeh, G.A., Talebi, A.F., Saraf, A., Suradkar, A., Tabatabaei, M., Singh, P., 2020. Description of novel species of *Aliinostoc, Desikacharya* and *Desmonostoc* using a polyphasic approach. International Journal of Systematic and Evolutionary Microbiology, 70, 3413-26.
Komárek, J., 2016. A polyphasic approach for the taxonomy of cyanobacteria: principles and applications. European Journal of Phycology, 51, 346-53.
Koonin, E.V., Makarova, K.S., Aravind, L., 2002. Horizontal gene transfer in prokaryotes: quantification and classification. Annual Review of Microbiology, 55, 709-742.
Leikoski, N., Fewer, D.P., Sivonen, K., 2009. Widespread occurrence and lateral transfer of the cyanobactin biosynthesis gene cluster in cyanobacteria. Applied and Environmental Microbiology, 75, 853-7.
Neilan, B.A., Jacobs, D., Blackall, L.L., Hawkins, P.R., Cox, P.T., Goodman, A.E., 1997. rRNA sequences and evolutionary relationships among toxic and nontoxic cyanobacteria of the genus *Microcystis*. International Journal of Systematic and Evolutionary Microbiology, 47, 693-7.
Nguyen, L.T., Schmidt, H.A., Von Haeseler, A., Minh, B.Q., 2015. IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. Molecular Biology and Evolution, 32, 268-74.
Nowruzi, B., Blanco, S., 2019. In silico identification and evolutionary analysis of candidate genes involved in the biosynthesis methylproline genes in cyanobacteria strains of Iran. Phytochemistry Letters, 29, 199-211.
Nowruzi, B., Soares, F., 2021. *Alborzia kermanshahica* gen. nov., sp. nov. (*Chroococcales*, Cyanobacteria), isolated from paddy fields in Iran. International Journal of Systematic and Evolutionary Microbiology, 30, 004828.
Nowruzi, B., Lorenzi, A.S., 2021. Characterization of a potentially microcystin-producing *Fischerella* sp. isolated from Ajigol wetland of Iran. South African Journal of Botany, 137, 423-33.
Nowruzi, B., Shalygin, S., 2021. Multiple phylogenies reveal a true taxonomic position of *Dulcicalothrix alborzica* sp. nov.(Nostocales, Cyanobacteria). Fottea, 21, 235-46.
Nowruzi, B., Hutárová, L., Absalón, I.B., Liu, L., 2022. A new strain of *Neowestiellopsis* (*Hapalosiphonaceae*): first observation of toxic soil cyanobacteria from agricultural fields in Iran. BMC microbiology, 22, 1-3.
Piccin-Santos, V., Brandão, M.M., Bittencourt-Oliveira, M.D., 2014. Phylogenetic study of *Geitlerinema* and *Microcystis* (Cyanobacteria) using PC-IGS and 16S–23S ITS as markers: investigation of

horizontal gene transfer. Journal of Phycology, 50, 736-43.
Rantala, A., Fewer, D.P., Hisbergues, M., Rouhiainen, L., Vaitomaa, J., Börner, T., Sivonen, K., 2004. Phylogenetic evidence for the early evolution of microcystin synthesis. Proceedings of the National Academy of Sciences, 13, 568-73.
Rippka, R., Deruelles, J., Waterbury, JB., Herdman, M., Stanier, R.Y., 1979. Generic assignments, strain histories and properties of pure cultures of cyanobacteria. Microbiology, 111, 1-61.
Robertson, B.R., Tezuka, N., Watanabe, M.M., 2001. Phylogenetic analyses of *Synechococcus* strains (cyanobacteria) using sequences of 16S rDNA and part of the phycocyanin operon reveal multiple evolutionary lines and reflect phycobilin content. International Journal of Systematic and Evolutionary Microbiology, 51, 861-71.
Sanchis, D., Padilla, C., Del Campo, F.F., Quesada, A., Sanz-Alferez, S. 2005. Phylogenetic and morphological analyses of *Microcystis* strains (Cyanophyta/Cyanobacteria) from a Spanish water reservoir. Nova Hedwigia, 81,431-436.
Six, C., Thomas, JC., Garczarek, L., Ostrowski, M., Dufresne, A., Blot, N., Scanlan, D.J., Partensky, F., 2007. Diversity and evolution of phycobilisomes in marine *Synechococcus* spp.: a comparative genomics study. Genome Biology, 12, 1-22
Taton, A., Grubisic, S., Brambilla, E., De Wit, R., Wilmotte, A. 2003. Cyanobacterial diversity in natural and artificial microbial mats of Lake Fryxell (McMurdo Dry Valleys, Antarctica): a morphological and molecular approach. Applied Environmental Microbiology, 69, 5157–5169.
Teneva, I., Stoyanov, P., Mladenov, R., Dzhambazov, B., 2012. Molecular and phylogenetic characterization of two species of the genus *Nostoc* (Cyanobacteria) based on the cpcB-IGS-cpcA locus of the phycocyanin operon. Journal of Bioscience and Biotechnology, 1, 9-19.
Zhaxybayeva, O., Gogarten, J.P., Charlebois, R.L., Doolittle, W.F., Papke, R.T., 2006. Phylogenetic analyses of cyanobacterial genomes: quantification of horizontal gene transfer events. Genome research, 16, 1099-108.
Zuker, M., 2013. Mfold web server for nucleic acid folding and hybridization prediction. Nucleic Acids Research, 31, 3406-15.
Yerrapragada, S., Siefert, J.L., Fox, G.E., 2009. Horizontal gene transfer in cyanobacterial signature genes. Horizontal Gene Transfer, 339-66.