THE PHYLOGENETIC RECONSTRUCTION OF BIOSYNTHESIZING GENES FROM CYANOBACTERIA ISOLATED FROM ZIARAT WATERFALL OF GOLESTAN PROVINCE

B. Nowruzi & S. Jafari Porzani

Cyanobacteria are anaerobic and photosynthesizing prokaryotes. In the past, the classification of cyanobacteria was only based on morphological characteristics. Today, advanced techniques such as molecular markers are used to achieve more accurate and reliable classification. In this research, sampling was done from the Ziarat waterfall in Golestan province. After cultivation and purification in solid culture medium BG-11, molecular identification of natural compounds' biosynthesizing genes (hassallidins synthetase (*has*N) and geosmin A (geoA) along with divergence analyses were performed using the amplification of the chloroplast genes RNA polymerase C1 (*rpo*C1). Then, the strain containing the above genes was identified by amplification of 16S rRNA and internal transcribed spacer (ITS) genes. Phylogenetic trees were built using the Maximum Likelihood method and the appropriate model with the help of a web server on the IQ-Tree server. The secondary structure of ITS was drawn in different parts of helix D1-D1′, D2, D3, tRNAIle, tRNAAla, BOX B, BOX A, and V3 using the Mfold program. The results showed that only the *Nodosilinea* sp. 1359 (Leptolyngbyaceae, Synechococcales) strain has the genes mentioned above. In addition, the results of calculating the *KA/KS* of *has*N and *geo*A genes and the phylogenetic incongruence of 16S rRNA and *rpo*C1 genes showed that natural selection by creating positive mutations has led to diversity in the studied strain. This study is among the first research conducted on the molecular phylogeny of cyanobacteria producing natural compounds in Ziarat waterfall.

Bahareh Nowruzi (correspondence<bahareh.nowruzi@srbiau.ac.ir>) & Samaneh Jafari Porzani, Department of Biotechnology, Faculty of Converging Sciences and Technologies, Islamic Azad University, Science and Research, Tehran, Iran.

بازسازی فیلوژنتیک ژن‌های بیوسنتزکننده سیانوباکتری‌های جدا شده از آبشار زیارت استان گلستان

بهاره نوروزی: استادیار گروه بیوتکنولوژی، دانشکده علوم و فناوری‌های همگرا، واحد علوم و تحقیقات، دانشگاه آزاد اسلامی، تهران

سمانه جعفری پرزانی: کارشناس ارشد گروه بیوتکنولوژی، دانشکده علوم و فناوری‌های همگرا، واحد علوم و تحقیقات، دانشگاه آزاد اسلامی، تهران، ایران

سیانوباکتری‌ها، پروکاریوت‌های بی‌هوازی و فتوسنتز کننده هستند. در گذشته ردهبندی سیانوباکتری‌ها تنها براساس صفات مورفولوژیکی استوار بود. اما امروزه از تکنیک‌های پیشرفته مانند نشانگرهای مولکولی برای دستیابی به ردهبندی دقیق‌تر و قابل اعتمادتر استفاده می‌شود. در این تحقیق نمونه برداری از آبشار زیارت استان گلستان انجام شد. پس از کشت و خالص‌سازی در محیط کشت جامد BG-11، شناسایی مولکولی ژن‌های بیوسنتزکننده ترکیبات طبیعی (*has*N، و *geo*A) به همراه آنالیزهای واگرایی با تکثیر ژن‌های *rpo*C1 انجام گردید. سپس سویه حاوی ژن‌های فوق با تکثیر ژن های 16S rRNA و ITS شناسایی گردید. درختان فیلوژنتیک با استفاده از روش بیشینه درست‌نمایی و مدل مناسب به کمک وب سرور IQ-Tree

ساخته شد. ساختار ثانویه ITS، در بخش‌های مختلف هلیکس D1-D1′، D2، D3، tRNAIle، tRNAAla، BOX B، BOX A و V3 به کمک برنامه M-fold رسم شد. نتایج نشان داد که تنها سویه 1359 .sp *Nodosilinea* دارای ژن‌های فوق بوده و متعلق به تیره Leptolyngbyaceae و راسته Synechococcales است. علاوه بر آن، نتایج حاصل از محاسبه میزان *KA/KS* ژن‌های *has*N و *geo*A و ناسازگاری داده‌های ژنتیکی 16S rRNA و *rpo* C1 نشان داد که انتخاب طبیعی با ایجاد جهش‌های مثبت منجر به ایجاد تنوع در سویه مورد مطالعه شده است. این مطالعه، جزو نخستین تحقیقات انجام شده روی فیلوژنی مولکولی سیانوباکتریوم تولیدکننده ترکیبات طبیعی در آبشار زیارت است.

INTRODUCTION

The process of classification plays a crucial role in comprehending and quantifying the vast array of living organisms. Nevertheless, the complexity of prokaryotes and phototrophic cyanobacteria, encompassing both simple unicellular forms and multicellular types, presents a challenge (Komárek 2016). In recent years, there has been a significant shift in the classification of cyanobacteria, particularly due to the adoption of molecular techniques and advanced methods of ultrastructural analysis (Nowruzi & Hutarova 2023; Nowruzi & Lorenzi 2023).

Historically, the classification of cyanobacteria relied on morphological and visual attributes, which were deemed inadequate and imprecise for the development of the contemporary cyanobacterial classification system. There is a divergence of perspectives regarding the quantification of cyanobacteria species owing to their considerable intricacy (Nowruzi & al. 2022; Nowruzi & Afshari 2023). There exists a divergence among researchers regarding the classification of species, with some opting to categorize them into various smaller classes, while others propose the creation of new classes to accommodate newly discovered species, primarily considering disparities in ecophysiology and morphology (Dvořák & al. 2015). Significant advancements have been achieved in the classification of cyanobacteria and the identification of genera through the utilization of the polyphasic approach and molecular sequences. The initial action was initiated in the year 2002, wherein the identification of the first genus was accomplished through the integration of morphological and molecular techniques (Abed & al. 2002).

Despite the extensive metabolic capabilities and fast growth of cyanobacteria, there has been a relative lack of research on their molecular phylogeny and evolutionary genetics. This knowledge gap poses challenges in identifying high-quality strains for bioactive compound production (Rastogi & al. 2010). The pharmaceutical industry has witnessed a significant increase in the utilization of cyanobacteria for the extraction and identification of various metabolites possessing novel medicinal properties.

These metabolites include but are not limited to anticancer agents, antibiotics, antiviral compounds, antifungal enzymes, and protease inhibitors. This recent development highlights the growing significance of cyanobacteria in the field of pharmaceutical research (Tan 2007). Hence, the genetic examination of cyanobacteria possessing valuable metabolic products represents a significant and substantial advancement in the characterization of cyanobacterial strains from a phylogenetic perspective.

The existence of polyphyletic genera leads to the need to use sufficient different molecular markers to study the closely related species. In some cases, the lack of resolution of traditional genetic markers, mainly the 16S rRNA, can lead to a need for the use of several different genes to identify the species belonging to these genera (Nowruzi & Lorenzi 2023).

Various protein-coding genetic sequences have been used for inferring phylogenies within cyanobacteria (*rpo*C1, *rpo*B, *gyr*B, *rbc*LX, *cpc*BA-IGS, and 16S-23S ITS), (Sciuto & al. 2012; Seo & Yokota 2003; Boyer & al. 2001; Premanandh & al. 2006; Cai & al. 2019; Neilan & al. 1995).

In the past, the use of different molecular markers such as *rpo*C1, *nif*D, *nif*H, *cpc*A, and *psb*A have helped to resolve the problem with closely related species. The *rpo*C1 gene, which encodes the β-subunit of RNA polymerase, is a more designating genetic marker among the closely related species (Nowruzi & Hutarova 2023). This marker was recently used in the study and description of the genus *Minunostoc* (Cai & al. 2019), *Neocylindrospermum variakineticum* (Tawong & al. 2022), and *Dulcicalothrix alborzica* (Nowruzi & Shalygin 2021).

The main cause of odor production in cyanobacteria is due to the presence of a compound called geosmin, which is stored in the cells and released in high concentration when the cells are destroyed. This organic compound causes many problems in drinking water and produces dangerous toxins (Nowruzi & Porzani 2021). Hassallidins are another group of antimicrobial peptidases that have antifungal activity against *Candida* spp. and *Cryptococcus neoformans* (Chlipala & al. 2011; Nowruzi 2022).

The primary objective of this study is to examine the molecular phylogeny of cyanobacterial strain _Nodosilinea_ sp. 1359 utilizing the 16S rRNA and ITS structural genes. Subsequently, the polymerase chain reaction (PCR) products of the _has_N, _geo_A, and _rpo_C1 genes were subjected to sequencing, followed by the construction of a phylogenetic tree. Furthermore, this is the first evidence of a potentially natural bioactive compounds -producing _Nodosilinea_ due to the presence of the _has_N and _geo_A genes.

MATERIALS AND METHODS
Sampling and cultivation and purification of cyanobacteria samples

Eleven water samples were obtained from Ziarat waterfall by employing plastic bottles. To accomplish this, the containers were meticulously and gradually rotated to allow for the collection of water, mud, and algae from the depth of 10 cm (Nowruzi & Shalygin 2021). The samples were transferred to sterile petri dishes containing 1.2% agar-solidified $BG11_0$ medium (Rippka & al. 1979) under aseptic conditions. After 20 days, the isolated colonies were selected, washed with sterilized deionized water, and transferred to 1 mL of fresh liquid $BG11_0$ medium. Intact filaments were plated again on solid $BG11_0$ medium by spread plate technique (Katoh & al. 2012) after 10-12 days of growth. The procedure was repeated until monocultures were obtained. Subsequently, the isolates were cultured in a 250 mL cotton stoppered Erlenmeyer flask containing 100 mL medium with pH adjusted to 7.2 at 28±2 ºC with periodic shaking (twice a day), illumination of ca. 50-55 μmoL photons $m^{-2}s^{-1}$, and a regime of 14:10 h light: dark cycle. After cyanobacterial isolation, the purification process was performed under sterile conditions. All the tests were performed triplets (Nowruzi & Becerra-Absalón 2022). An Olympus CX31RTS5 (Olympus, Tokyo, Japan) was utilized for morphological observation of the culture (Miscoe & al. 2016).

DNA isolation and PCR amplification

Genomic DNA was isolated from 16-18 days old log phase cultures using the Power Soil^{TM} DNA Isolation Kit (MO BIO Laboratories, Inc., Carlsbad, CA, USA), according to the manufacturer´s instructions. PCR amplification of the 16S rRNA, Internal Transcribed Spacer (ITS), the chloroplast genes RNA polymerase C1 (_rpo_C1), hassallidins synthetase (_has_N) and geosmin A (_geo_A) gene's regions were performed in a Bio-Rad iCycler (Bio-Rad, USA) thermocycler using the oligonucleotide primer sets listed in Table 1.

Reactions were made using 10-20 ηg DNA template, 0.5 μM of each primer, 1.5 mM $MgCl_2$, 200 μM dNTPs, 1U/μL Taq DNA polymerase (Invitrogen, Carlsbad, CA, USA) and ultrapure water to a 25-μL final volume. The thermocycler was programmed according to the program shown in Table 1. All PCR products stained with SYBR$^{\circledR}$ Gold Nucleic Acid Gel Stain (Invitrogen, Carlsbad, CA, EUA) were checked by electrophoresis on 1% agarose gels (SeaPlaque® GTG, Cambrex Corporation) using standard protocols, and recorded. Subsequently, amplicons were purified using the GeneClean$^{\circledR}$ Turbo kit (Qbiogene, MP Biomedicals) prior to sequencing.

DNA sequencing and analysis

DNA sequencing was carried out using the refined PCR products and the BigDye$^{\circledR}$ Terminator v3.1 Cycle Sequencing Kit (Applied Biosystems, Life Technologies), following the manufacturer's instructions. The target sequences were bidirectionally sequenced, and each set of sequencing data was obtained from at least three independent sequencing reactions. The pair sequenced fragments were assembled into contigs using the BioEdit Sequence Alignment Editor version 7 (Hall 1999). Only bases with standard quality (Phred score> 20) were considered. Then, pairwise comparisons with other cyanobacterial sequences available in the GenBank from NCBI database were made using the BLAST tool (http://www.ncbi.nlm.nih.gov/BLAST) and validated information. Especially for the 16S rRNA gene sequences, the similarity search was also performed using the EZtaxon database (www.eztaxon.org) (Kim & al. 2012) with validated cyanobacterial strains only.

Phylogenetic analysis

The 16S ribosomal RNA gene and 16S-23S ribosomal RNA internal transcribed spacer, _rpo_C1, _has_N and _geo_A sequences obtained in this study as well as the best hit sequences (>94 % identity) retrieved from GenBank were aligned using MAFFT version 7 (https://mafft.cbrc.jp/alignment/server/) (Katoh & Standley 2013). Maximum likelihood analyses were performed using partial 16S rRNA gene sequences containing a maximum of 661 characters including nucleotides and indels. Then maximum likelihood phylogenetic trees were inferred in IQ-Tree (multicore v1.5.5), (Nguyen & al. 2015) with 573, 19, 227 and 197 sequences compared during phylogenetic analysis of 16S ribosomal RNA gene and 16S-23S ribosomal RNA intergenic spacer, _rpo_C1, _has_N and _geo_A genes respectively. Optimum models were used as suggested (BIC criterion) after employing the model test implemented in IQ-tree (Table 2). Tree robustness was estimated with bootstrap percentages using 100 standard bootstrap searches and 10,000 ultrafast bootstrap to evaluate branch supports. FigTree v.1.2.2 was used for tree visualization.

In order to calculate the percentage of genetic similarity between strains, p-distances analysis was used. First, the sequences were aligned with each other using the Clustal W program. The aligned file was opened with the MegaX program and after selecting the best model (TN93+G), pairwise distances were performed. In the obtained Excel file, in order to obtain the percentage of genetic similarity, the number obtained was subtracted from the number one and multiplied by the number one hundred (Nowruzi & Becerra-Absalón 2022).

Table 1: Target genes and oligonucleotide primers used in this study.

Target gene	Primers with their sequences (5´-3´)	PCR condition	Reference
16S rRNA	PA (5'-AGAGTTTGATCCTGGCTCAG-3') B23S (5'-CTTCGCCTCTGTGTGCCTAGGT-3')	94˚C, 3 min 30 x (94˚C, 30 s; 55˚C, 40 s; 72˚C, 1.30 min) 72˚C, 3 min 4˚C, ∞	(Taton & al. 2003)
16S-23S rRNA ITS	ITS-F (5'-TGTACACACCGCCCGTC-3') ITS-R (5'-CTCTGTGTGCCTAGGTATCC-3')		(Iteman & al. 2000)
*rpo*C1	5'-TGGGGHGAAAGNACAYTNCCTAA-3' 5'-GCAAANCGTCCNCCATCYAAYTGBA-3'	95°C, 3 min; 30_ (94°C, 30 sec; 56°C, 30 sec; 72°C, 1 min); and 72°C, 10 min	(Rantala & al. 2004)
*geo*A	geoA-297f (5'-RTCGAGTACATCGAGATGCG-3') geoA-552r (5'-CGBGAGGTGAGGAYGTCGTT-3')	5 min at 94 °C, followed by 38 cycles of 45 s at 94 °C, 45 s at 55 °C, and 2 min at 72 °C, followed by a final elongation step of 10 min at 72 °C.	(Shardlow 2021)
*has*N	GTA GAT GCG GTG CCA TTG AC GAC TAC CAC TGA TTG CTT CCA C	94 ºC for 3 min, 36 cycles of 94 ºC for 30 s, 50 ºC for 30 s and 72 ºC for 1 min, and a final extension at 72 ºC for 10 min.	(Vestola & al. 2014)

Table 2. Accession numbers of sequence data deposited in the DNA Data Bank of Japan.

Target gene	Strain	Nucleotide ID	Length of region (bp)	Number of amino acids	Tree model
16S ribosomal RNA gene and 16S-23S ribosomal RNA intergenic spacer	B1359	OR513468	661	-	TVM+F+I+G4
*rpo*C1	B1359	OR518654	675	225	TVMe+I+G4+F
*has*N	B1359	OR518655	972	324	TIM2+F+I+G4
*geo*A	B1359	OR518656	1596	532	SYM+G4

16S-23S rRNA ITS secondary structure analysis

The Sequence corresponding to the D1-D1' helix, D2, D3, BOX B, BOX A and D4 regions of the 16S-23S ITS of the studied strain were characterized according to Johansen & al. (2011). Comparison of the ITS secondary structures of the studied strain and reference strains were generated using the M-fold web server (version 2.3) under ideal conditions of untangled loop fix and the temperature set to default (37° C).

Sequence divergence of hassallidins and geosmin genes

The number of nonsynonymous substitutions per nonsynonymous site (KA) and the number of synonymous substitutions per synonymous site (KS), using MEGA 6. 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. Moreover, identifications of the predicted amino acid activated by NRPS A module and the probably name of the compound of *has*N gene was performed using software located at http://www.tigr.org/jravel/nrps (Rantala & al. 2004).

RESULTS

Morphology and macroscopic samples

Microscopic inspection of the materials allowed to the identification of filamentous cyanobacterial colonies with presence of heterocytes and akinetes. The results from optical microscope revealed that the strain is stringy and strapped. The morphology of studied strain was consistent with *Nodosilinea* sp.,

(Leptolyngbyaceae, Synechococcales), because it presented typical characteristics of the genus, with cells longer than wide and akinetes that form chains. The filaments are straight and slightly curved. Filaments are long (up to more than 100 cells long), straight or gently curved under high light, uniseriate, without false branching. Trichomes slightly motile and constricted at the cross-walls (Fig. 1).

Phylogenetic and p-distances analyses of 16S rDNA gene

A 16S rDNA fragment was sequenced and aligned with other 573 nucleotide sequences of cyanobacteria obtained from GenBank for phylogenetic analysis. The tree of maximum Likelihood (ML) phylogenetic analysis is shown in Fig. 2. The studied strain (*Nodosilinea* sp. 1359) has been placed together with other *Nodosilinea* species with bootstrap support 99.7 percent.

According to the phylogenetic tree we have also compared the 16S rRNA and ITS *p*-distances of our strain with related genera namely (*Nodosilinea* sp. 1359, *Nodosilinea* sp. SLM0509, *Nodosilinea* sp. SLM0611, *Nodosilinea* sp. SLM0106, *Nodosilinea* sp. SLM0611, *Nodosilinea* sp. SLM0106 and *Nodosilinea* sp. SLM0509) respectively.

Results showed that *Nodosilinea* sp. 1359 shared a 16S rRNA sequence similarity of 99.088% with *Nodosilinea* sp. SLM0509 (OK030540_1_1125-1439), 99.7933% with *Nodosilinea* sp. SLM0611 (OK030535_1_989-1359), 99. 01% with (OK030535_1_731-970) *Nodosilinea* sp. SLM0611, 100% with (OK030534_1_1140-1389) *Nodosilinea* sp. SLM0106, 100% with (OK030534_1_882-1121) *Nodosilinea* sp. SLM0106 and 100% with (OK030540_1_867-1106) *Nodosilinea* sp. SLM0509 (Table 3).

Fig. 1. Microscopic images of the strain under the microscope (Bars= 10 µm). H, Heterocytes; A, Akinetes.

98.8/76
Nodosilinea (22 OTUs)
85.1/44 — MT946554_1_1324-1907_Nodosilinea_cf__signiensis_NN-2-1-EE_
76.5/56
Nodosilinea (8 OTUs)
78/72
80.4/96 — OK030540_1_1125-1439_Nodosilinea_sp__SLM0509_
OK030535_1_989-1359_Nodosilinea_sp__SLM0611_
100/100 — OK030534_1_1140-1389_Nodosilinea_sp__SLM0106_
93.5/63
Nodosilinea sp. 1359
99.7/87 — OK030540_1_867-1106_Nodosilinea_sp__SLM0509_
86.4/76
99.7/87 — OK030535_1_731-970_Nodosilinea_sp__SLM0611_
OK030534_1_882-1121_Nodosilinea_sp__SLM0106_
74.5/73
MT946557_1_1324-1900_Nodosilinea_sp__OD-2-1-CH_
10.8/29 — OM732236_1_1353-1939_Nodosilinea_sp__BACA0422_
99.9/100 — MK502141_1_992-1571_Nodosilinea_sp__2AD401_Zammit_
82.1/80 — MK502140_1_1309-1887_Nodosilinea_sp__SA1303_Zammit_
93.8/100 — OK030539_1_957-1517_Nodosilinea_sp__SLM0112_
OK030538_1_969-1529_Nodosilinea_sp__SLM0111_
91.4/94
MF348321_1_1346-1922_Nodosilinea_ramsarensis_KH-S_S2_6_clone_cl4_
100/100 — MF348318_1_1325-1901_Nodosilinea_ramsarensis_KH-S_S2_6_clone_cl1_
95.7/87 — MF348319_1_1331-1907_Nodosilinea_ramsarensis_KH-S_S2_6_clone_cl2_
MF348320_1_1325-1901_Nodosilinea_ramsarensis_KH-S_S2_6_clone_cl3_
KM438183_1_1303-1879_Nodosilinea_sp__IkpPStn44_
99.6/97
ON897688_1_987-1583_Nodosilinea_sp__KPABG_T4_strain_T4__KPABG_
85.1/44
Nodosilinea (7 OTUs)
56.1/31 29.6/51 — MH688851_1_1334-1916_Nodosilinea_cf__epilithica_Us-2-1_
43.7/46 — HM018678_1_1034-1616_Nodosilinea_sp__FI2-2HA2_
66.6/62 — KJ939095_1_1034-1488_Nodosilinea_sp__CMT-3FSIN-NPC22B_
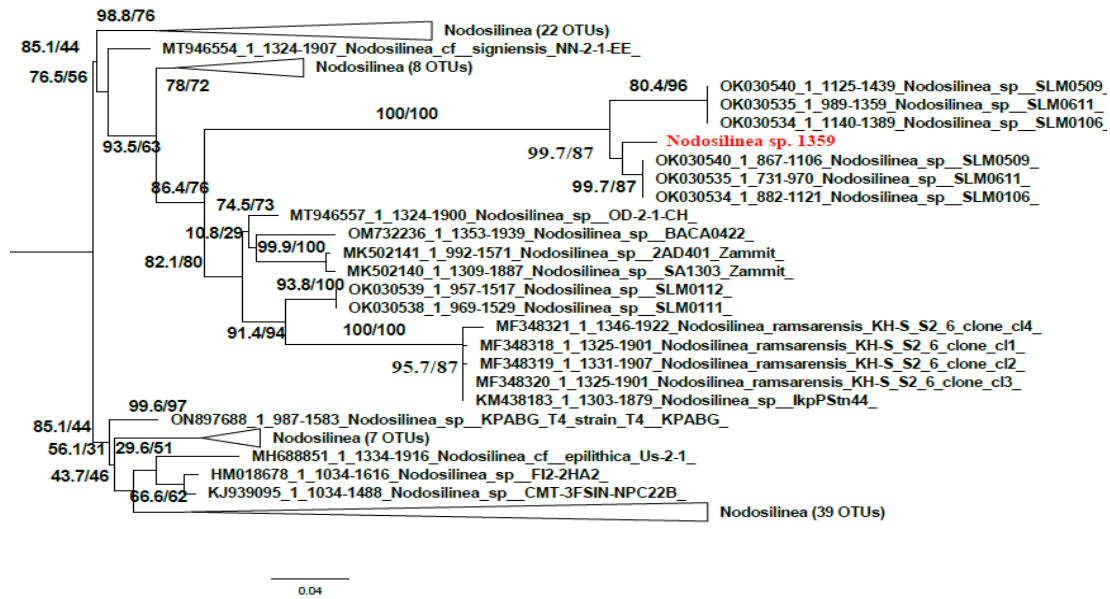Nodosilinea (39 OTUs)

0.04

Fig. 2. Phylogenetic relationships among *Nodosilinea* sp. 1359 (in red) and related cyanobacteria based on 16S rDNA sequences (573 bp). Numbers near nodes indicate standard bootstrap support (%)/ultrafast bootstrap support (%) for ML analyses.

Table 3. 16S rRNA gene sequence similarity matrix of *Nodosilinea* sp. 1359 and related taxa.

	(OK030540-1125-1439) *Nodosilinea* sp. SLM0509	(OK030535-989-1359) *Nodosilinea* sp. SLM0611	(OK030535-731-970) *Nodosilinea* sp. SLM0611	(OK030534-1140-1389) *Nodosilinea* sp. SLM0106	(OK030534_882-1121) *Nodosilinea* sp. SLM0106	(OK030540_1_867-1106) *Nodosilinea* sp. SLM0509	*Nodosilinea* sp. 1359
(OK030540_1_1125-1439) *Nodosilinea* sp. SLM0509							
(OK030535_1_989-1359) *Nodosilinea* sp. SLM0611	99.1206						
(OK030535_1_731-970) *Nodosilinea* sp. SLM0611	99.088	99.7933					
(OK030534_1_1140-1389) *Nodosilinea* sp. SLM0106	99.088	99.7933	99.99				
(OK030534_1_882-1121) *Nodosilinea* sp. SLM0106	99.088	99.7933	99.01	100			
(OK030540_1_867-1106) *Nodosilinea* sp. SLM0509	99.088	99.7933	99.01	100	100		
Nodosilinea sp. 1359	99.088	99.7933	99.01	100	100	100	100

Phylogenetic analyses and sequence divergence of hassallidins and geosmin genes

*has*N and *geo*A genes did not supported the position of *Nodosilinea* sp. 1359, because the presence of these genes in this genus has not been investigated so far. In the *has*N and *geo*A phylogenetic trees, *Nodosilinea* sp. 1359 was placed together with the *Leptolyngbya* (WP_017289430) and *Leptolyngbya* sp. A2 (AJP00077) with a phylogenetic similarity of 90.4 and 100% respectively. *Leptolyngbya* also belongs to family Leptolyngbyaceae (Figs. 3 & 4).

Results of identifications of the predicted amino acid activated by NRPS *A* module and the probable name of the compound of *has*N gene showed that *Nodosilinea* sp. 1359 screened, possessed only 1 A-domain, while the remaining possessed 2 or more. These A-domains can result in the production of one compound, or multiple independent compounds. The signature sequence, the name of the compound, and the predicted amino acid of *Nodosilinea* sp. 1359 was not

the same with other strains. Therefore, clustering of these NRPS sequences, illustrates a lack of taxonomic affiliations between cyanobacteria and A-domains (Table 4). For example, few instances of close relationship between the taxonomic status and the predicted compound were evident.

To determine whether the hassallidins and geosmin synthetase genes are under positive or negative selection pressure, we compared the number of nonsynonymous substitutions per nonsynonymous site (KA) to the number of synonymous substitutions per synonymous site (KS).

The KA/KS ratio was well greater than 1 in both genes. The ratio was $KA/KS=0.1762633772/0.0885981721=1.989469681169641>1$ and $KA/KS=0.07952 / 0.03174 =2.50>1$ for hassallidins and geosmin synthetase genes respectively. The KA/KS ratio greater than one, is a common method for identifying positive selection in molecular evolutionary studies and leads to diversity in strain.
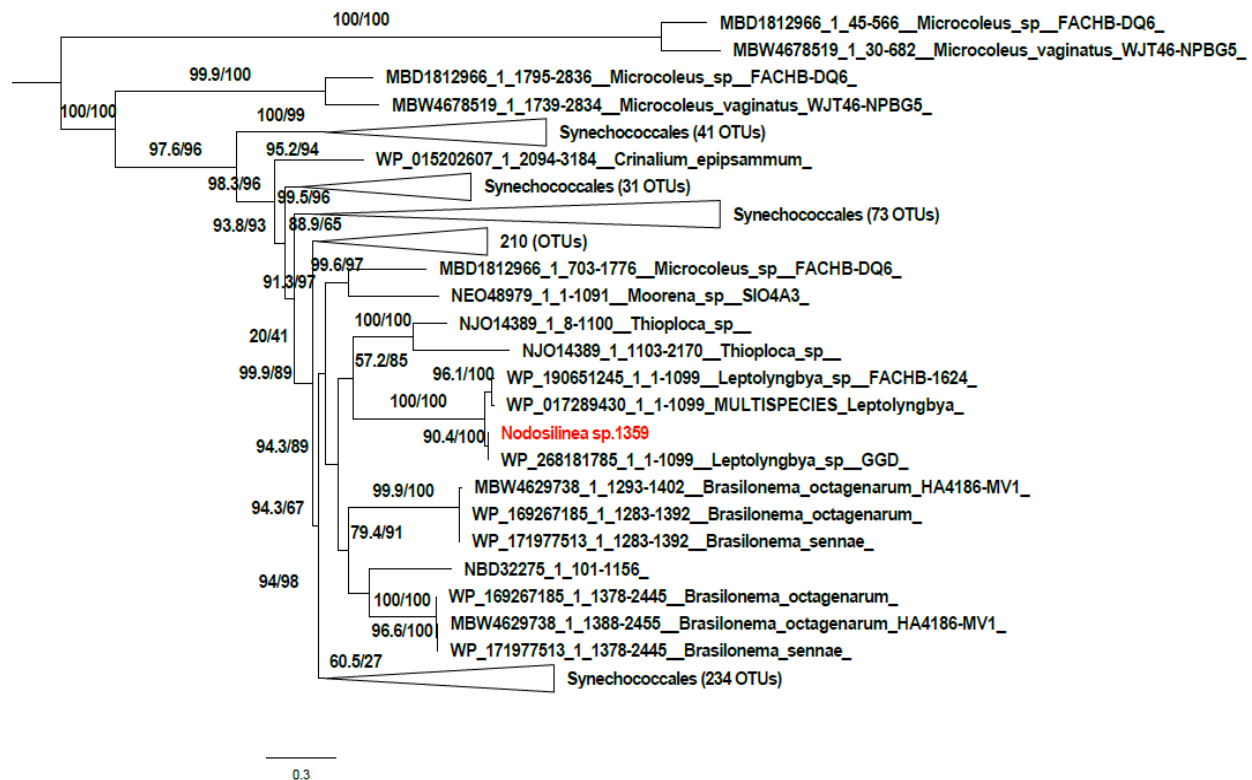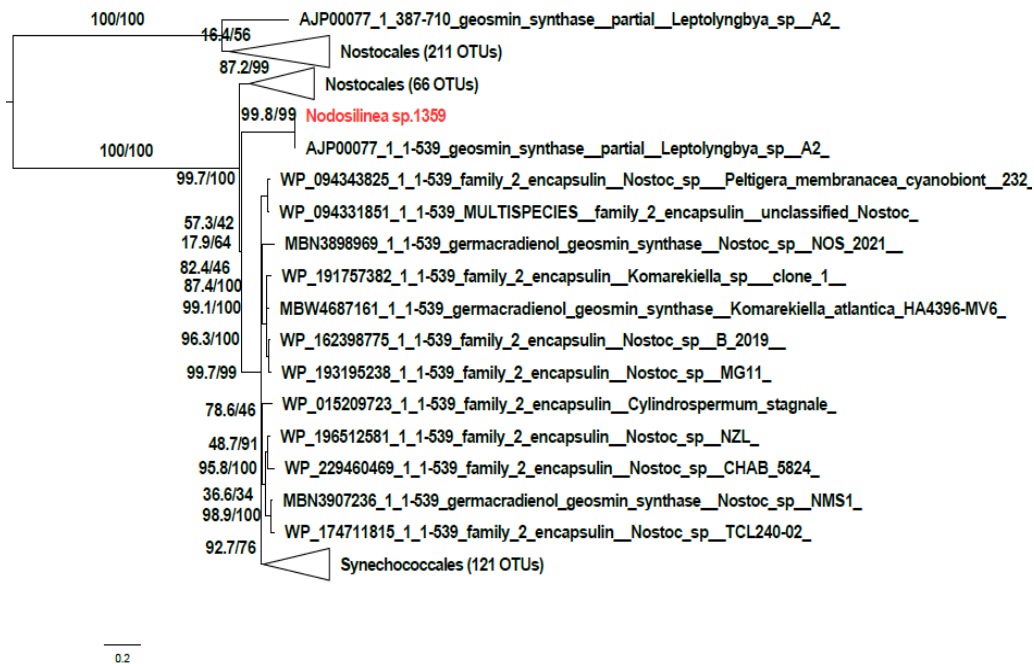

Fig. 3. Phylogenetic relationships among *Nodosilinea* sp. 1359 (in red) and related cyanobacteria based on *has*N gene sequences. Numbers near nodes indicate standard bootstrap support (%)/ultrafast bootstrap support (%) for ML analyses.

Table 4: Analysis of the translated NRPS A-domain proteins showed the signature sequence and substrate-binding pocket of the metabolite biosynthesis pathway.

Compounds	Predicted amino acid	Signatures sequences	Number of domains	Strain	Accession no.
Chloroeremomycin synthetase	CepB-M1-HPG	DVFNLGLI	A-domain 1	*Nodosilinea* sp. 1359	OR513468
	NO HIT	DILXICLI	A-domain 1	*Thioploca* sp. NJO14389	NJO14389
Anabaenopeptilide synthetase D	AdpD-M1-Ile	DAFFLGVT	A-domain 2		
Anabaenopeptilide synthetase D	NO HIT	DILXICLI	A-domain 1	*Thioploca* sp. NJO14389	NJO14389
	AdpD-M1-Ile	DAFFLGVT	A-domain 2		
Chloroeremomycin synthetase	CepB-M1-HPG	DVFNLGLI	A-domain 1	*Leptolyngbya* sp. FACHB-1624	WP_190651245
Chloroeremomycin synthetase	CepB-M1-HPG	DVFNLGLI	A-domain 1	*Leptolyngbya* sp. WP_017289430.1	WP_017289430
Chloroeremomycin synthetase	CepB-M1-HPG	DVFNLGLI	A-domain 1	*Leptolyngbya* sp. GGD	WP_268181785

Fig. 4. Phylogenetic relationships among *Nodosilinea* sp. 1359 (in red) and related cyanobacteria based on geoN gene sequences. Numbers near nodes indicate standard bootstrap support (%)/ultrafast bootstrap support (%) for ML analyses.

Incongruence between the 16S rRNA and *rpo*C1

The sequence divergence between the taxa included in this analysis from the 16S rRNA gene and *rpo*C1 data set were comparable. The high degree of incongruence between the 16S rRNA and *rpo*C1 data set are consistent with diversity in strain (Fig. 5). In fact, the rpoC1 data was able to distinguish the genera, and the genus *Nostoc* was separated from the genera *Leptolyngbya* and *Synechococcus*. Moreover *Nodosilinea* sp. 1359 has been placed together with *Synechococcus* with bootstrap support 99 percent (Fig. 5).
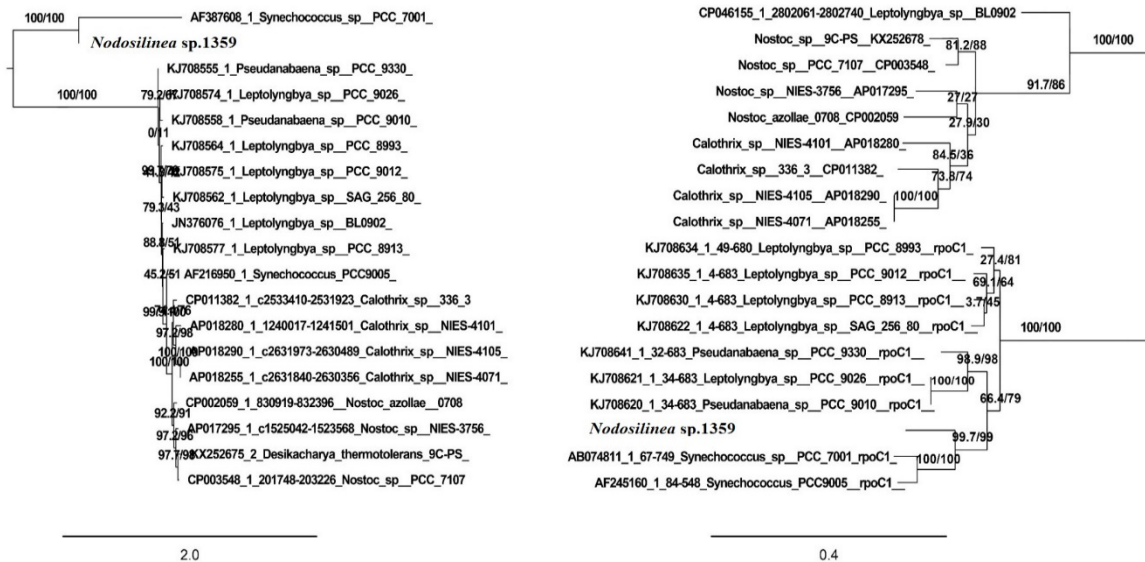
Fig 5. Phylogenetic incongruence of rpoC1 and 16srRNA gene drawn by Iq tree web server (16srRNA tree on the left and rpoC1 tree on the right).

16S-23S rRNA ITS secondary structure

Six reference sequences were used to search for ITS secondary structure. According to Johansen & al. (2011), nine different areas (D1-D1' helix, D2, D3, BOX B, BOX A, D4 and V3 helix) were found in the ITS secondary structure of their 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 (Figs. 6 & 7). The lengths of D1-D1' helices were shown to be the same for all studied strains. Concerning Box-B, lengths varied from 55 nt in (OK030540_1_1125-1439) *Nodosilinea* sp. SLM0509, (OK030535_1_989-1359) *Nodosilinea* sp. SLM0611, (OK030534_1_1140-1389) *Nodosilinea* sp. SLM0106, (OK030534_1_882-1121) *Nodosilinea* sp. SLM0106, (OK030535_1_731-970) *Nodosilinea* sp. SLM0611 and (OK030540_1_867-1106) to 44 nt in *Nodosilinea* sp. 1359.

Moreover, V3 helix was only found in (OK030535_1_989-1359) *Nodosilinea* sp. SLM0611 and (OK030535_1_731-970) *Nodosilinea* sp. SLM0611 and the studied strain. The V3 helix was varied in terms of length and shape among the studied strain and reference strains (Fig. 8), (Tables 5-7).

DISCUSSION

It is estimated that a large percentage of cyanobacterial strains worldwide have been wrongly identified and diagnosed. The use of DNA sequences allows to identify the strains and infer the phylogeny of organisms (Komárek 2016). The results of this research show rRNA sequences can be used to distinguish unicellular and filamentous species. Likewise, the study of the nucleotide sequence of rRNA genes showed that genera that are grouped in a consensus family and have endospore-producing genes are grouped in related clusters or branches in phylogenetic trees (Prabha & Singh 2019). The results of their research showed that the strains that had the same producer genes were grouped in a clade, which is exactly in accordance with the results of our research. However, the results of the clustering of domain A of the *hassallidin* gene illustrates a lack of taxonomic affiliations between cyanobacteria and A-domains (Table 4). For example, few instances of close relationship between the taxonomic status and the predicted compound were evident.

So far, molecular investigations that have been carried out in the field of identifying genes that produce natural compounds, on the genus *Nodosilinea*, are very limited (Vestola & al. 2014; Cerón-Vivas & al. 2023; Tawong & al. 2022). The results of phylogenetic trees based on hassallidin and geosmin genes were clustered with the genera in the Leptolyngbyaceae family.
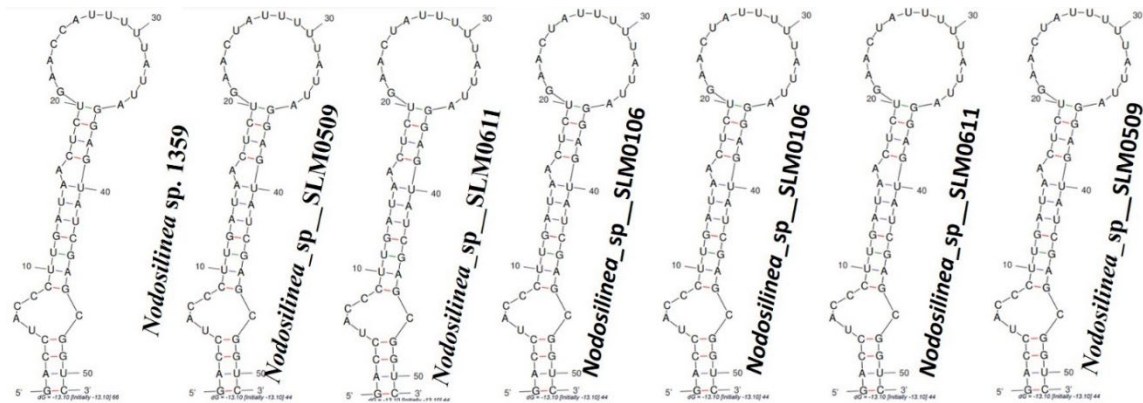
Fig. 6. Comparison of secondary structures of the D1–D1′ helixes from 16S–23S intergenic spacers between *Nodosilinea* sp. 1359 and reference strains.
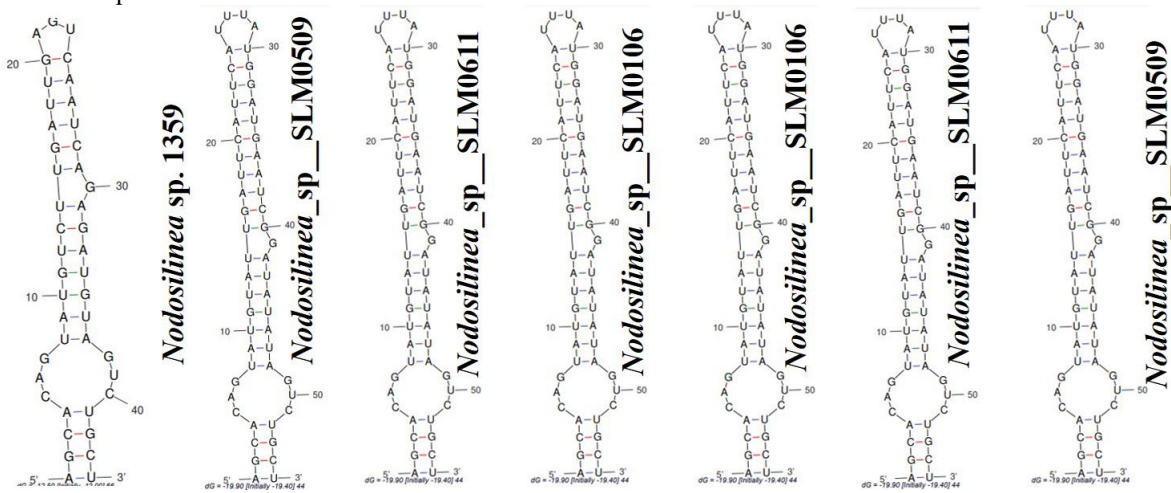

Fig. 7. Comparison of secondary structure of the Box B helixes from 16S–23S intergenic spacers between *Nodosilinea* sp. 1359 and the reference strains.
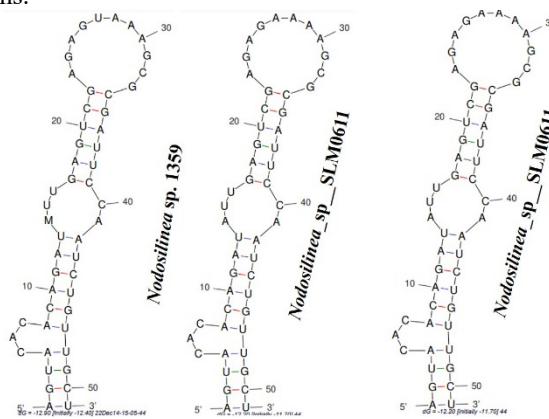

Fig. 8. Comparison of secondary structure of the V3 helixes from 16S–23S intergenic spacers between *Nodosilinea* sp. 1359 and reference strains.

Table 5. Comparison of the nucleotide's length of the ITS regions of *Nodosilinea* sp.1359 and reference strains.

Studied strain and reference strains	D1-D' helix	spacer+D 2+ spacer	D3	D3 + spacer	trRNAIle gene	TrRNAAla gene	BOX B	BOX A	D 4 + spacer	V3	D4
Nodosilinea sp.1359	51	42	4	71	74	73	44	11	24	67	9
(OK030540_1_1125-1439) *Nodosilinea* sp. SLM0509	51	42	4	30	74	73	55	11	23	-	9
(OK030535_1_989-1359) *Nodosilinea* sp. SLM0611	51	42	4	30	74	73	55	11	26	51	9
(OK030534_1_1140-1389) *Nodosilinea* sp. SLM0106	51	42	4	30	74	73	55	-	-	-	-
(OK030534_1_882-1121) *Nodosilinea* sp. SLM0106	51	42	4	30	74	73	55	-	-	-	-
(OK030535_1_731-970) *Nodosilinea* sp. SLM0611	51	42	4	30	74	73	55	11	26	51	9
(OK030540_1_867-1106) *Nodosilinea* sp. SLM0509	51	42	4	30	74	73	55	11	26	-	9

Table 6. Comparison of secondary structure of 16S-23S rRNA (D1-D1' helix and Box-B helix) between the *Nodosilinea* sp. 1359 and reference strains.

Studied Strain and reference strains	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)
	Nucleotide	Loop	Loop	Nucleotide	Loop	Nucleotide
Nodosilinea sp.1359	17	1	1	8	5	10
(OK030540_1_1125-1439) *Nodosilinea* sp. SLM0509	17	1	1	8	6	10
(OK030535_1_989-1359) *Nodosilinea* sp. SLM0611	17	1	1	8	6	10
(OK030534_1_1140-1389) *Nodosilinea* sp. SLM0106	17	1	1	8	6	10
(OK030534_1_882-1121) *Nodosilinea* sp. SLM0106	17	1	1	8	6	10
(OK030535_1_731-970) *Nodosilinea* sp. SLM0611	17	1	1	8	6	10
(OK030540_1_867-1106) *Nodosilinea* sp. SLM0509	17	1	1	8	6	10

ITS sequence is much more diverse than gene sequence of 16S rRNA and is widely used to distinguish different species within the same genus. For example, Řeháková & al. (2007) used ITS secondary structure variation to compare *Nostoc commune* and *Nostoc punctiforme* (Řeháková & al. 2007). Also, Bohunická & al. (2015), used ITS gene sequence and Box-B and V3 helices to identify four species of the genus *Roholtiella* (Bohunická & al. 2015). In the current research, various regions in ITS were used to further separate the strain producing natural compounds from other strains, while the D1-D1 helix did not show any difference among the reference strains, B Box region contained significant differences between the studied strain and the other referenced sequences.

Table 7. Comparison of secondary structure of 16S-23S rRNA (V3' helix) between the *Nodosilinea* sp.1359 with reference strains.

Studied Strain and reference strains	V3' helix			
	Terminal Bilateral Bulge (A)	Bilateral Bulge (B)	Unilateral Bulge (C)	Basal Clamp (D)
	Nucleotide	Loop	Loop	Nucleotide
Nodosilinea sp. 1359	13	1	1	8
(OK030540_1_1125-1439) *Nodosilinea* sp. SLM0509	-	-	-	-
(OK030535_1_989-1359) *Nodosilinea* sp. SLM0611	13	1	1	8
(OK030534_1_1140-1389) *Nodosilinea* sp. SLM0106	-	-	-	-

Rivandi & al. (2021) investigated the morphology and phylogeny of the toxic strain purified from Lavasan Lake water, using 16S rRNA and ITS gene markers (Rivandi & al. 2021). Based on the results obtained from the phylogeny tree drawn based on the 16S rRNA gene, the non-toxic strains with the toxic strain of *Anabaena* sp. B3 (CCC B3), were not clustered within a clade. The results of ITS gene analysis using Mfold showed that the most important difference between the strain producing natural compounds *Nodosilinea* sp. 1359 and other similar strains is the number of Box-B nucleotides and the presence of V3 region.

Analysis of the ITS region in seven *Oculatella* species showed that it is similar to most of the Pseudanabaenales and has at least two operons, one with two tRNAs and the other with no tRNAs. Helices in different operons have similar structures. In this study, all studied strains had both tRNAs.

The results of *p*-distances analyses of 16S rDNA gene of studied strain showed that *Nodosilinea* sp. 1359 represented a 16S rRNA sequence similarity of more than 99.088 % with the references sequences. According to the Yarza & al. 2014, identities <98.7% are considered strong evidence for considering compared strains to be in different species, while identities <94.5% are considered to be strong evidence for considering compared strains to be in different genera (Yarza & al. 2014). Therefore, the studied strain does not belong to a new species or genus.

The construction and comparison of phylogenetic trees are perhaps the best ways to assess the contribution of horizontal gene transfer to the evolutionary history of a gene family. Incongruence is taken to indicate a role for horizontal gene transfer, whereas congruence is consistent with descent through common ancestry. In this study we compared two data sets comprised of *rpo*C1 and 16srRNA genes. These sequences were analyzed and tested for congruence. Based on the findings of this investigation, it can be inferred that the incongruity in the phylogenetic placement of the *rpo*C1 gene and 16srRNA in phylogenetic trees are likely attributable to the horizontal transfer of cyanobacterial biosynthetic genes and also different phylogenetic signals because of their different evolutionary pathways.

The present study has determined that the utilization of functional and structural protected sequences in the genome of cyanobacteria, such as hassallidins synthetase (*has*N) and geosmin A (geoA), *rpo*c1, ITS, and 16SrRNA sequences, can be valuable in elucidating genomic relationships and discerning strain types. The study demonstrates significant utility, highlighting the necessity of employing multiple genes in conjunction for the purpose of identification. This method can serve as a diagnostic tool for differentiating cyanobacteria and identifying them across various biological, climatic, and geographical contexts. Furthermore, by conducting an analysis to identify the potential presence of cyanobacteria in the ponds of Ziarat waterfall, it becomes feasible to contemplate suitable strategies pertaining to the optimal water treatment approach.

REFERENCES

Abed, R.M., Garcia-Pichel, F. & Hernández-Mariné, M. 2002: Polyphasic characterization of benthic, moderately halophilic, moderately thermophilic cyanobacteria with very thin trichomes and the proposal of Halomicronema excentricum gen. nov., sp. nov. -Arch. Microbiol. 177: 361–370. DOI: 10.1007/s00203-001-0390-2

Bohunická, M., Pietrasiak, N., Johansen, J., Berrendero-Gomez, E., Hauer, T., Gaysina, L. & Lukešová, A. 2015: Roholtiella, gen. nov. (Nostocales, Cyanobacteria)- a tapering and branching member of the Nostocaceae (Cyanobacteria). -Phytotaxa 197(2): 84–103. DOI: 10.11646/phytotaxa.197.2.2

Boyer, S.L., Flechtner, V.R. & Johansen, J.R. 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(6): 1057–1069. DOI: 10.1093/oxfordjournals.molbev.a003877

Cai, F., Li, X., Yang, Y., Jia, N., Huo, D. & Li, R. 2019: Compactonostoc shennongjiaensis gen. & sp. nov.(Nostocales, Cyanobacteria) from a wet rocky wall in China. -Phycologia 58(2): 200–210. DOI: 10.1080/00318884.2018.1541270

Dvořák, P., Poulíčková, A., Hašler, P., Belli, M., Casamatta D. A. & Papini, A. 2015: Species concepts and speciation factors in cyanobacteria, with connection to the problems of diversity and classification. Biodivers. -Conserv. 24: 739–757. DOI: 10.1007/s10531-015-0888-6

Chlipala, G. E., Mo, S. & Orjala, J. 2011: Chemodiversity in freshwater and terrestrial cyanobacteria- a source for drug discovery. -Curr. Drug Targets. 12(11): 1654–1673. DOI: 10.2174/138945011798109455

Hall, T.A. 1999: BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. -Nucleic acids symposium series 41: 95–98. London.

Iteman, I., Rippka, R., Tandeau de Marsac, N. & Herdman, M. 2000: Comparison of conserved structural and regulatory domains within divergent 16S rRNA–23S rRNA spacer sequences of cyanobacteria. -Microbiology. 146(6): 1275–1286. DOI: 10.1099/00221287-146-6-1275

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(3): 283. DOI: 10.1127/0029-5035/2011/0092-0283

Katoh, H., Furukawa, J., Tomita-Yokotani, K. & Nishi, Y. 2012: Isolation and purification of an axenic diazotrophic drought-tolerant cyanobacterium, Nostoc commune, from natural cyanobacterial crusts and its utilization for field research on soils polluted with radioisotopes. -Biochim Biophys Acta Bioenerg. 1817(8): 1499–1505. DOI: 10.1016/j.bbabio.2012.02.039

Katoh, K. & Standley, D.M. 2013: MAFFT multiple sequence alignment software version 7: improvements in performance and usability. -Mol. Biol. Evol. 30(4): 772–780. DOI: 10.1093/molbev/mst010

Kim, O.S., Cho, Y.J., Lee, K., Yoon, S.H., Kim, M., Na, H., Park, S.C., Jeon, Y. S., Lee, J.H. & Yi, H. 2012: Introducing EzTaxon-e: a prokaryotic 16S rRNA gene sequence database with phylotypes that represent uncultured species. -IJSEM. 62(3): 716–721. DOI: 10.1099/ijs.0.038075-0

Komárek, J. 2016: A polyphasic approach for the taxonomy of cyanobacteria: principles and applications. -Eur. J. Phycol. 51(3): 346–353. DOI: 10.1080/09670262.2016.1163738

Miscoe, L.H., Johansen, J.R., Kociolek, J.P., Lowe, R.L., Vaccarino, M.A., Pietrasiak, N. & Sherwood, A.R. 2016: The diatom flora and cyanobacteria from caves on Kauai, Hawaii. -Acta Bot. Hung. 58: 3–4.

Neilan, B.A., Jacobs, D. & Goodman, A.E. 1995: Genetic diversity and phylogeny of toxic cyanobacteria determined by DNA polymorphisms within the phycocyanin locus. -AEM. 61(11): 3875–3883. DOI: 10.1128/aem.61.11.3875-3883.1995

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. -Mol. Biol. Evol. 32(1): 268–274. DOI: 10.1093/molbev/msu300

Nowruzi, B. 2022: Cyanobacteria Natural Products as Sources for Future Directions in Antibiotic Drug Discovery. -Recent Advances and New Perspectives 1: 1–19. DOI: 10.5772/intechopen.106364

Nowruzi, B. & Afshari, G. 2023: In silico Analysis of Molecular Phylogeny of Genes Involved in the Synthesis of Bioactive Compounds in Cyanobacteria Strains Located in Tehran Cascade. -JJCMB. 14(1). DOI: 10.5812/jjcmb-132400

Nowruzi, B. & Becerra-Absalón, I. 2022: A Novel Potentially Toxic Cyanobacterial Species From the Genus Desmonostoc, Desmonostoc Alborizicum sp. nov., Iisolated From a Water Supply System of IranA. -Res. Sq. 1: 1–19.

Nowruzi, B., Hutarova, L. 2023: Structural and functional genes, and highly repetitive sequences commonly used in the phylogeny and species concept of the phylum Cyanobacteria. -Cryptogamie, Algologie. 44(3): 59–84. DOI: 10.5252/cryptogamiealogie2023v44a3

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 Microbiol. 22(1): 1–13. DOI: 10.1186/s12866-022-02525-x

Nowruzi, B. & Lorenzi, A.S. 2023: Molecular phylogeny of two Aliinostoc isolates from a paddy field. -Plant Syst. Evol. 309(2): 11. DOI: 10.1007/s00606-023-01848-0

Nowruzi, B. & Porzani, S.J. 2021: Toxic compounds produced by cyanobacteria belonging to several species of the order Nostocales: A review. -J. Appl. Toxicol. 41(4): 510–548. DOI: 10.1002/jat.4088

Nowruzi, B. & Shalygin, S. 2021: Multiple phylogenies reveal a true taxonomic position of Dulcicalothrix alborzica sp. nov. (Nostocales ,Cyanobacteria). -Fottea 21(2): 235–246.

Prabha, R. & Singh, D.P. 2019: Cyanobacterial phylogenetic analysis based on phylogenomics approaches render evolutionary diversification and adaptation: an overview of representative orders. -3 Biotech. 9: 1–16. DOI:10.1007/s13205-019-1635-6

Premanandh, J., Priya, B., Teneva, I., Dzhambazov, B., Prabaharan, D. & Uma, L. 2006: Molecular characterization of marine cyanobacteria from the Indian subcontinent deduced from sequence analysis of the phycocyanin operon (cpcB-IGS-cpcA) and 16S-23S ITS region. -J. Microbiol. 44(6): 607–616.

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. -PNAS. 101(2): 568–573. DOI: 10.1073/pnas.0304489101

Rastogi, R.P., Singh, S.P., Häder, D.P. & Sinha, R.P. 2010: Detection of reactive oxygen species (ROS) by the oxidant-sensing probe 2′, 7′-dichlorodihydrofluorescein diacetate in the cyanobacterium Anabaena variabilis PCC 7937. -BBRC. 397(3): 603–607. DOI: 10.1016/j.bbrc.2010.06.006

Řeháková, K., Johansen, J.R., Casamatta, D.A., Xuesong, L. & Vincent, J. 2007: Morphological and molecular characterization of selected desert soil cyanobacteria: three species new to science including Mojavia pulchra gen. et sp. nov. -Phycologia 46: 481–502.

Rippka, R., Deruelles, J., Waterbury, J.B., Herdman, M. & Stanier, R.Y. 1979: Generic assignments, strain histories and properties of pure cultures of cyanobacteria. -Microbiology. 111(1): 1–61. DOI: 10.1099/00221287-111-1-1

Rivandi, M., Nowruzi, B. & Fahimi, H. 2021: Molecular phylogenetic study of toxic cyanobacterium Anabaena sp. strain B3 isolated from Lavasan Lake, Tehran (Iran). -Rostaniha 22(1): 120–133. DOI: 10.22092/BOTANY.2021.355307.1259

Sciuto, K., Andreoli, C., Rascio, N., La Rocca, N. & Moro, I. 2012: Polyphasic approach and typification of selected Phormidium strains (Cyanobacteria). -Cladistics 28(4): 357–374. DOI: 10.1111/j.1096-0031.2011.00386.x.

Seo, P.S. & Yokota, A. 2003: The phylogenetic relationships of cyanobacteria inferred from 16S rRNA, gyrB, rpoC1 and rpoD1 gene sequences. -J. Gen. Appl. Microbiol. 49(3): 191–203. DOI: 10.2323/jgam.49.191

Shardlow, T. 2021: Identification and characterization of toxic cyanobacteria in two forested maritime watersheds in North America, University of Waterloo. -UWSpace. DOI: hdl.handle.net/10012/17055

Tan, L.T. 2007: Bioactive natural products from marine cyanobacteria for drug discovery. -Phytochemistry. 68(7): 954–979. DOI: 10.1016/j.phytochem.2007.01.012

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. -AEM. 69(9): 5157–5169 . DOI: 10.1128/AEM.69.9.5157-5169.2003

Tawong, W., Pongcharoen, P. & Saijuntha, W. 2022: "Neocylindrospermum variakineticum gen. & sp. nov.(Nostocales, Cyanobacteria), a novel genus separated from Cylindrospermum using a polyphasic method. -Phycologia. 61(6): 653–668. DOI: 10.1080/00318884.2022.2130829

Vestola, J., Shishido, T.K., Jokela, J., Fewer, D.P., Aitio, O., Permi, P., Wahlsten, M., Wang, H., Rouhiainen, L. & Sivonen, K. 2014: Hassallidins, antifungal glycolipopeptides, are widespread among cyanobacteria and are the end-product of a nonribosomal pathway. -Proceedings of the National Academy of Sciences 111(18): E1909–E1917. DOI: 10.1073/pnas.1320913111

Yarza, P., Yilmaz, P., Pruesse, E., Glöckner, F.O., Ludwig, W., Schleifer, K.H., Whitman, W.B., Euzéby, J., Amann, R. & Rosselló-Móra, R. 2014: Uniting the classification of cultured and uncultured bacteria and archaea using 16S rRNA gene sequences. -Nat. Rev. Microbiol. 12(9): 63645. DOI: 10.1038/nrmicro3330