 doi:10.20944/preprints202202.0189.v1

Type of the Paper: Article

Genomic Insights on the Functional Capabilities of the Cyano-Sphere of Edible Andean *Nostoc* Macrocolonies (Llayta)

Claudia Vilo[1], Alexandra Galetovic[1], Qunfeng Dong[2], Benito Gómez-Silva [1]*

[1] Laboratory of Biochemistry, Biomedical Department, Health Sciences Faculty and Centre for Biotechnology and Bioengineering (CeBiB), Universidad de Antofagasta, Antofagasta, Chile; claudiavilo@gmail.com; alexandra.galetovic@uantof.cl; benito.gomez@uantof.cl

[2] Center for Biomedical Informatics, Department of Medicine, Stritch School of Medicine, Loyola University of Chicago, Maywood, Illinois, USA Affiliation 2; qdong@luc.edu

* Correspondence: benito.gomez@uantof.cl

Abstract: Cyanobacteria biomasses are sources of secondary metabolites and nutritious ingredients such as vitamins, essential amino acids, and unsaturated fatty acids. Biochemical composition, presence of cyanotoxins and contaminants are major concerns to be addressed on such edible biomasses. Macrocolonies of a filamentous diazotrophic *Nostoc* species known as Llayta are found at Andean wetlands and consumed since pre-Columbian times in South America. Its biochemical composition has been previously conducted to assess their nutritious quality and cyanotoxicity. Macrocolonies of filamentous cyanobacteria are niches for colonization by diverse microorganisms; however, the Llayta microcolonies cyanosphere is unknown. Based on a culture-independent approach, we report the identification of members of the resilient microflora associated with Llayta trichomes after Gentamicin treatments. We have also reconstructed the genomes of the Llayta macrocolony-forming *Nostoc* sp. cyanobacterium (6,781,030 bp; GC content of 41.2%) and the genomes of five dominant bacteria genera (*Mesorhizobium*, *Microvirga*, *Paracoccus*, *Aquimonas*, and *Blastomonas*). The detection of genes and genes clusters involved in primary and secondary metabolism is described. Our results provide new insights on the metabolic capabilities and biotechnological potential of the Andean *Nostoc* cyanobacterium, and the ecological role and adaptive strategies of microorganisms living under extreme environmental conditions at the Andean wetlands.

Keywords: cyanosphere, cyanobacteria, Llayta, *Nostoc* macrocolonies, metagenomics, microbiome

1. Introduction

Cyanobacterial and microalgal biomasses have been consumed for centuries [1-8]. Dunaliella, Chlorella and Arthrospira, in Africa and North America, and *Nostoc*, in Asia and South America, are documented sources of food containing essential amino acids, vitamins, and polyunsaturated fatty acids, but also carotenoids, phycobiliproteins, and secondary metabolites with diverse biotechnological applications [8-13].

The Atacama Desert is considered the driest and oldest desert on Earth, with climatic, oceanic, latitudinal, and geomorphological factors that account for its existence [14,15]. The abundance and diversity of life forms in this hyper-arid desert are limited by high solar insolation and the absence of reliable liquid water sources [15]. Comparatively, precipitations at the Andes Range are several hundred times higher than those observed at the dry core of the Atacama Desert, with an evident and positive impact on the flora and microflora at the Andean wetlands [16].

The Andean biome has been historically acknowledged by the indigenous Andean cultures as source of forage, food, and ethnomedicine [1-3,8,15,14,17]. The consumption of macrocolonies of a filamentous diazotrophic *Nostoc* species, known by the vernacular name of Llayta, is a centenary Andean alimentary practice that can be traced back to pre-

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 15 February 2022 doi:10.20944/preprints202202.0189.v1

Columbian times and it is an example of a neglected natural food resource [8]. Free-living Llayta macrocolonies are collected at wetlands at the Andes highlands over 3,000 m of altitude. They are sun-dried and sold for human consumption as a dry, dark green, leaf-like biomass in food markets in Arica and Iquique, in northern Chile, and Tacna, in southern Peru [8,17]. Besides the absence of epidemiological evidence against its centuries-old consumption, ethnographic, nutritional quality, and genomic studies on cyanotoxin biosynthesis studies support the notion that Llayta is a safe, natural food ingredient. In addition, the cyanobacterium *Nostoc* sp. Llayta has been identified as a non-cyanotoxin strain, the strain that forms edible Llayta macrocolonies in nature [8,17,18].

The natural life cycle of primary producers may involve mutualistic, antagonistic, or commensal interactions with cyanophages, fungi, and heterotrophic bacteria. Positive or negative outcomes from such interactions have been previously discussed in *Nostoc* macrocolonies [18,19] and reported in Andean *Nostoc* macrocolonies [20-22]. However, studies on identifying members of the bacterial community associated with edible Llayta macrocolonies have been a pending issue. Previous attempts to purify axenic Llayta filaments were unsuccessful due to the presence of high heterotrophic bacterial titers. Since isolated trichomes from the mucilaginous inner matrix of Llayta colonies showed the presence of an associated microflora [23] we decided to perform an antibiotic treatment to the filaments to isolate the *Nostoc* cyanobacterium. However, there was still a resilient bacterial community after gentamicin treatment. Then, metagenomics analyses were applied to identify the antibiotic-resilient bacterial community on Gentamycin-treated Llayta filaments.

Here we report metagenomics-based identification of prominent members of the Gentamicin-resilient microflora associated with isolated Llayta trichomes, the genome reconstructions, identification and annotation of functional genes, and insight on metabolic capabilities for the Andean *Nostoc* sp. Llayta cyanobacterium.

2. Materials and Methods

Isolation of Nostoc sp. Llayta filaments. Dried Llayta colonies were obtained from the major farmers market in Arica, Chile, during 2015 and maintained in the original plastic bag until used. Dry colonies were suspended in sterile deionized water for rehydration for 24 h. To isolate filaments, aliquots (5 mL) were transferred to liquid, nitrogen-free Arnon medium and cultured at 30°C under white fluorescent light (μE 180 m-2 s-1), continuous agitation (200 rpm), and aeration enriched in 1% v/v CO_2 [17]. Culture aliquots were spread onto agar plates, observed under a stereo microscope, and isolated filaments *Nostoc* sp. Llayta were picked and transferred to a fresh growth medium.

Antibiotic treatment. Aliquots (10-15 mL) of cultured filaments of *Nostoc* sp. Llayta were collected at the exponential growth phase, washed with fresh medium, and the filaments were recovered as a pellet by centrifugation at 4,000 x g, for 5 min, at room temperature. The cell pellet was suspended in 20 mL of fresh Arnon medium containing Gentamicin (1 mg/mL; Sigma Aldrich, Chile), Casamino acids (1.6 mg/mL; Sigma Aldrich, Chile), and D-glucose (0.8 mg/mL; Sigma Aldrich, Chile). The suspension was incubated for 48 h in darkness at 30°C and 120 rpm. The filament suspension was recovered by centrifugation, extensively washed with fresh culture medium, and grown in liquid Arnon medium. The biomass recovered by centrifugation was used to extract total genomic DNA.

DNA extraction. Total genomic DNA was extracted from the filament pellets with Ultra Clean Microbial DNA isolation kit (MoBio), following the manufacturer´s instructions. DNA quality was evaluated by electrophoresis in 0.8 % agarose gel and quantified photometrically at 260 nm.

DNA sequencing and analysis. The Llayta metagenome was sequenced via MiSeq sequencing technology using shotgun paired-end libraries, with an average insert size of 250 bp. Reads had an average length of 300 bp, with good quality scores, as evaluated by the FastQC program (version 0.10.0). The sequencing produced a total of 17,137,246 reads

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 15 February 2022 doi:10.20944/preprints202202.0189.v1

that were submitted to the Rapid Annotation using Subsystems Technology for Metagenomes (MG-RAST) web server [24] for taxonomic and functional assignment using default parameters. In addition, metagenomic assembly was done using MEGAHIT assembler v.1.2.9 [25], and binning was conducted using the PATRIC web server [26]. Secondary metabolites were searched with PRISM [27] and AntiSMASH [28] software. Sequencing reads are available at the Sequence Read Archive (SRA) with accession number SRR17916224. The Whole Genome Shotgun projects have been deposited at DDBJ/ENA/GenBank under the accession JAKOMP000000000, JAKOMQ000000000, JAKOMR000000000, JAKOMS000000000, JAKOMT000000000, and JAKOMU000000000.

3. Results

Metagenomics studies were conducted on filaments of *Nostoc* macrocolonies (Llayta) to identify the accompanying antibiotic-resilient bacteria and the cyanobacterium *Nostoc* sp. Llayta and to gain an insight into its functional metabolic capabilities.

3.1. Microbial diversity in Llayta trichomes

Taxonomic assignments obtained from metagenomics analyses of shotgun metagenome showed that the microbiota from gentamicin-treated filaments was dominated by bacteria (99%), while representatives of the domains Archaea (0.2%) and Eukarya (0.1%) were present at a much lower extent. Predominant bacteria in Llayta trichomes belong to the phylum *Proteobacteria* (82%) with *Xanthomonas* (38%), *Stenotrophomonas* (15%), and *Methylobacterium* (9.3%) as dominant genera (Fig. 1). The cyanobacterial phylum (16%) was dominated by the genera *Nostoc* (40%), *Anabaena* (26%), and *Nodularia* (21%). Archaea representatives were present at a lesser extent (0.2%). The genera *Xanthomonas* (38%) and *Nostoc* (40%) were the most abundant among bacteria and cyanobacteria (Fig. 1).

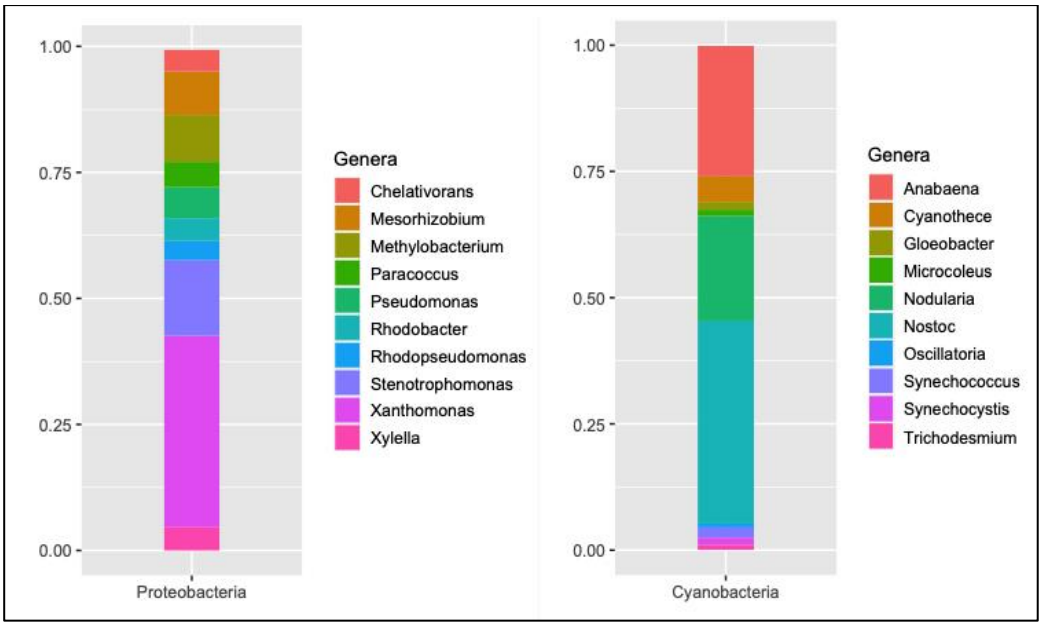

Figure 1. Genus-level relative abundance of the ten dominant *Proteobacteria* and *Cyanobacteria* found in the Gentamycin-treated Llayta filaments.

Two nitrogen-fixing alpha-proteobacteria from the genera *Mesorhizobium* and *Microvirga*, one nitrogen denitrifying bacterium belonging to the genus *Paracoccus*, one gamma-proteobacteria representative of the genus *Aquimonas*, and one aerobic photoheterotrophic alpha-proteobacteria from the genus *Blastomonas*, were identified among the most abundant microbial species accompanying the filaments from Llayta macrocolonies after

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 15 February 2022 doi:10.20944/preprints202202.0189.v1

antibiotic treatment. The corresponding phylogenetic analyses by the Maximum Likelihood method based on 16S rRNA gene sequences are shown in Supplementary Fig. S1. From the metagenomic sequences, metagenomes of the major microbial components of the Llayta microbiome were assembled, and summary statistics of metagenome assembly is shown in Table 1.

Table 1. Statistics for assembling and binning of the reconstructed metagenomes of prominent microbial members of Gentamycin-treated Llayta filaments.

Genus	Sequence size	Contigs number	N50 value	Completeness (%)
Nostoc sp. Llayta	6,781,030	244	50,369	100
Paracoccus	4,116,456	107	139,894	100
Microvirga	3,719,434	70	273,976	98.4
Mesorhizobium	4,745,523	165	427,465	91.3
Blastomonas	3,312,218	641	7,513	98.1
Aquimonas	4,464,829	80	155,856	98.3

3.2. Functional capabilities of the Llayta microbiome

Using the Subsystem annotation at MG-RAST, the Llayta metagenome contained many genes with putative metabolic capabilities related to adaptation to extreme environmental conditions: fatty acid metabolism (2.9%), dormancy and sporulation (0.14%), vitamin, prosthetic groups, pigments biosynthesis (5.5%), DNA metabolism (5.3%), stress response (3%), and resistance to toxic compounds (2%) (Fig. 2).

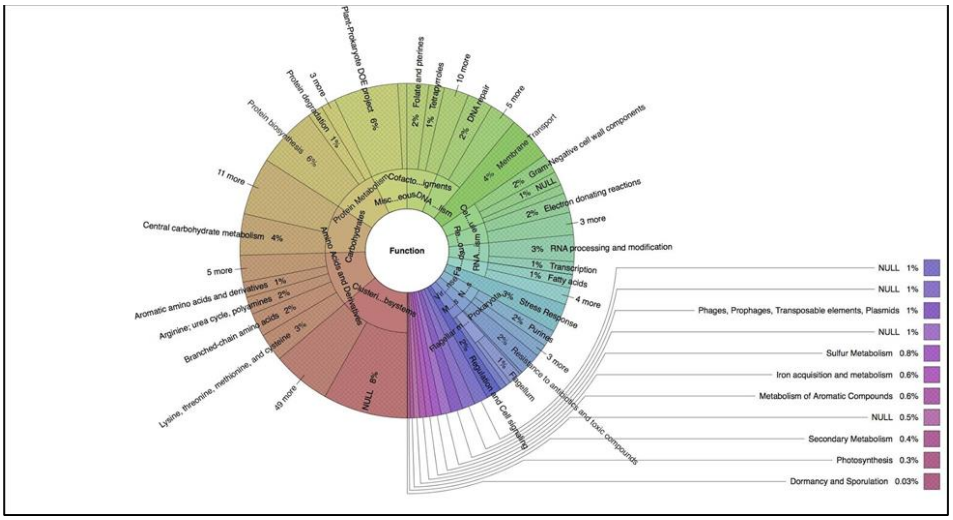

Figure 2. Functional analysis of the metagenome of Gentamycin-treated Llayta filaments.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 15 February 2022 doi:10.20944/preprints202202.0189.v1

3.3. Genome assembly and phylogeny of Nostoc sp. Llayta

Given the abundant *Nostoc* sequences found in the metagenomic analysis of gentamicin-treated Llayta filaments, the *Nostoc* genome was reconstructed using 244 contigs with an N50 of 50,369, which rendered a total genome size of 6,781,030 bp with a GC content of 41.2%. Phylogenetic analyses of concatenated marker genes (*16S rRNA*, *recA*, *dnaJ*, and *gyrB*) confirmed that the cyanobacterium-forming Llayta macrocolony belonged to the *Nostoc* genus (Fig. 3). However, no clear placement to a clade was observed, preventing assignation to a particular *Nostoc* species.

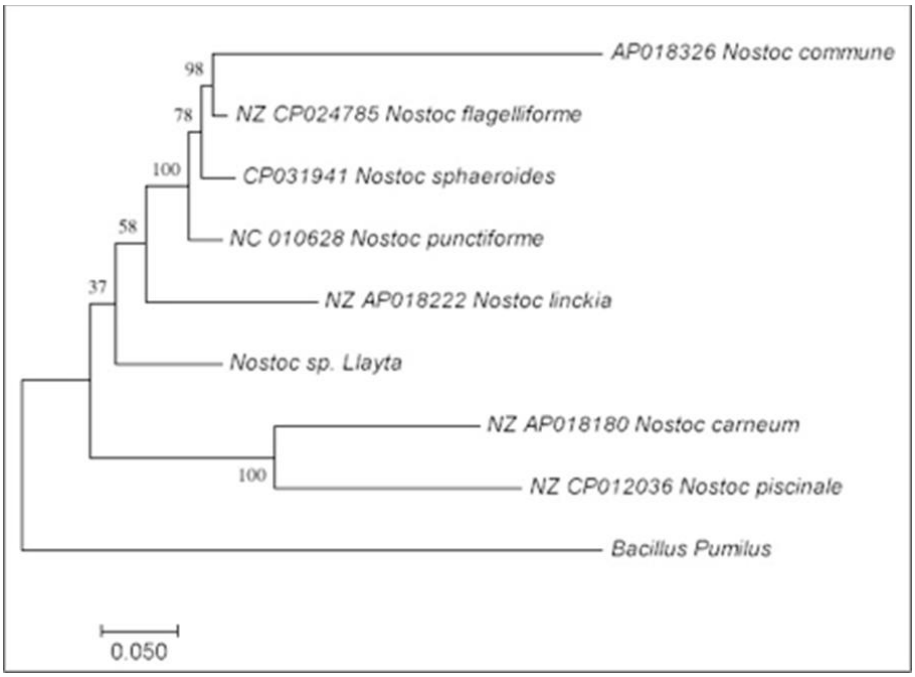

Figure 3. Phylogenetic analysis by Maximum Likelihood method based on concatenated marked gene sequences (*16S rRNA*, *recA*, *dnaJ*, and *gyrB*). The horizontal bar at the figure's base represents 0.05 substitutions per nucleotide site. The percentage of trees in which the associated taxa clustered together is shown next to the branches, using a bootstrap of 1000. Evolutionary analysis was conducted in MEGA X.

3.4 Functional capabilities of Nostoc sp. Llayta

The reconstructed genome of *Nostoc* sp. Llayta was analyzed to identify sequences of functional genes and gene clusters coding for enzymes involved in primary and secondary metabolism. The cyanobacterium genomic capabilities included sequences related to the production of vitamins, prosthetic groups, and pigments (5.5%); DNA metabolism (5.3%); fatty acids, lipids, and isoprenoids (2.9%); dormancy and sporulation (0.14%); multiple gene copies associated to Photosystems I and II proteins, genes and pathways associated to DNA repair metabolism, and genes for cell division and cell cycle processes. Gene sequences involved in chemotaxis and motility were absent.

The in-silico genomic analysis demonstrated the presence of conserved lycopene cyclase CruA/CruP genes, the complete *mysABCD* and *scyABCDEF* (and accompanied regulator and kinase sensor) gene clusters, involved in carotenoids biosynthesis, mycosporine-like amino acids (MAAs), and scytonemin biosynthesis, respectively (Table 2). These putative capabilities are intimately related to protection against photooxidative damages triggered by high-UV solar insolation prevailing at the Andes plateau wetlands [15,29].

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 15 February 2022 doi:10.20944/preprints202202.0189.v1

Table 2. *Nostoc* sp. Llayta open reading frames annotations and best Blast hit for MysABC and ScyABCDEF cluster genes

ORF	Description based on Subsystem annotations	Accession	Identity (%)
Orf 3147	Demethyl 4-deoxygadusol synthase MysA	WP_069074324.1	93
Orf 3148	O-methyltransferase MysB	BBC27542.1	85
Orf 3149	ATP-grasp ligase forming mycosporine-glycine, MysC	BBC27543.1	82
Orf 4592	scytonemin biosynthesis protein ScyA	WP_206262883.1	83
Orf 4593	tryptophan dehydrogenase ScyB	WP_086764771.1	84

Bioinformatic analyses rendered two copies of gene *cphA1* coding for a putative non-ribosomal cyanophycin synthetase in the reconstructed genome of *Nostoc* sp. Llayta. Sequence alignment of *Nostoc* sp. Llayta cyanophycin synthetase to Nostocales CphA1 (NCBI accession number: MBE9052550.1 and MBE9053887.1) showed an 85% identity. The cyanobacterium genome also contained three copies of gene *cphB* coding for a putative cyanophycin-degrading cyanophycinase and the cyanophycinase sequence from *Nostoc* sp. Llayta showed an 83% identity after alignment with CphA1 from Nostocales (NCBI accession number: AHJ26983.1 and WP_096726859.1).

Nostoc sp. Llayta genome contained sequences associated with the gene LanM, involved in the biosynthesis of a family of ribosomally-synthesized and post-translationally modified peptides. This sequence and the CCG motif for Zn binding sites at its C-terminal showed a 70% identity after alignments with the LanM sequence of cyanobacteria bacterium UBA11372 (NCBI accession number HAX78725.1). Fig. 4 shows the domains for dehydratase and cyclization activities and three cysteine residues associated with zinc-binding sites, one cysteine residue alone, and the CCG conserved motif from the Llayta cyanobacterium.

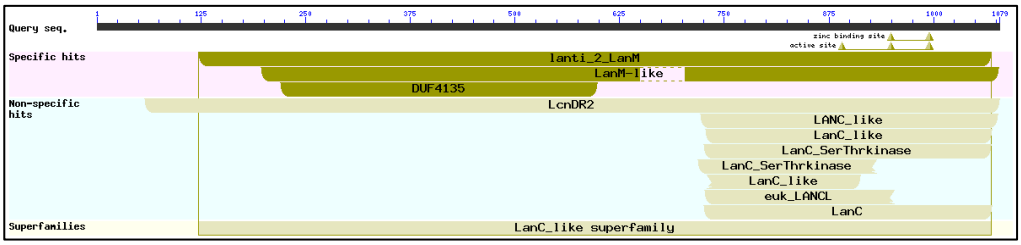

Figure 4. Domain characterization of lantibiotics biosynthesis enzyme using the NCBI Conserved Domain Database.

The reconstructed genome of *Nostoc* sp. Llayta contains sequences for all genes coding for lasso peptide biosynthesis: a lasso peptide precursor lasso P (49% identity to Nostocales, cyanobacterium accession number MBE9052603.1), a lasso C isopeptide bond-forming cyclase (74% identity to Nostocales, cyanobacterium MBE9052604.1), a lasso peptide biosynthesis B2 protein (73% identity to Nostocales, cyanobacterium accession number MBE9052605.1), and lasso peptide biosynthesis B1 PqqD family protein (86% identity to Nostocales, cyanobacterium accession number MBE9052606.1). Fig 5 depicts the organization for the lasso peptide gene cluster in the genome of the Llayta cyanobacterium, resembling that from Actinobacteria, and includes the putative primary structure of the cleaved lasso peptide from *Nostoc* sp. Llayta, the location of the lasso structure with a glycine residue at the peptide N-terminal and bulky residues at its C-terminal, and a comparison of conserved sequences among several microbial genera.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 15 February 2022 doi:10.20944/preprints202202.0189.v1

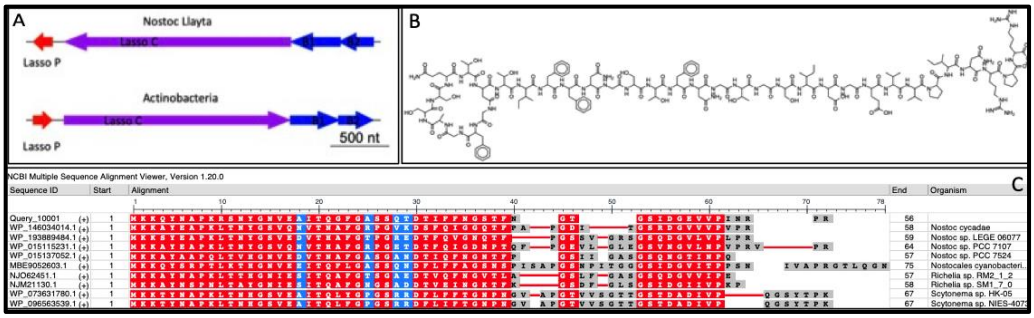

Figure 5. Characterization of the lasso peptides from the genome of the Llayta cyanobacterium. (A) Lasso peptide gene cluster organization in the genome of the Llayta cyanobacterium, resembling that from Actinobacteria. (B) Primary chemical structure of the cleaved Lasso P, made of 36 residues. (C) Protein sequence alignment of Lasso P peptide against NCBI non-redundant database.

The genome of *Nostoc* sp. Llayta also has putative biosynthetic pathways for capreomycidine peptides, from non-ribosomal peptide synthases (NRPS), and microviridin peptides, from ribosomally synthesized and post-translationally modified peptides (RiPPs). Capreomycidine synthase gene from *Nostoc* sp. Llayta showed an 82% identity with capreomycidine synthase gene (sequence ID: WP_104904844.1) of a *Nostoc* sp. cyanobiont of *Lobaria pulmonaria* and 42.23% identity with capreomycidine hydroxylase (sequence ID: WP_051702360) from *Streptomyces vinaceous*. Also, microviridin precursor peptide sequence in the *Nostoc* sp. Llayta genome showed an 83% identity with the microviridin family tricyclic proteinase inhibitor gene of *Nostoc* sp. (sequence ID: WP_224090535.1). The capreomycidine and microviridin biosynthetic pathway included arginine hydrolase and condensation, adenylation, thiolation, epimerization, and reductase enzymes.

CRISPR-Cas (Clustered Regularly Interspaced Short Palindromic Repeats-CRISPR-associated) operons are defensive systems found in archaea and bacteria [29-31] and were in *Nostoc* sp. Llayta genome. One of them included Cmr2-Cmr3-Cmr4-Cmr5-Cmr6 and Cas1-Cas2 operons flanked by CRISPR arrays; the second system showed the Csc3-Csc2-Csc1-Cas6-Cas4-Cas1-Cas2 genes and CRISPR array, and the third comprised Cmr2-Cmr3-Cmr4-Cmr5-Cmr6 genes flanked by CRISPR array (Table 3). Therefore, the CRISPR-Cas system operons in *Nostoc* sp. Llayta, according to the observed organization, would indeed confer immunity defense against mobile elements.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 15 February 2022 doi:10.20944/preprints202202.0189.v1

Table 3. *Nostoc* sp. Llayta open reading frame annotations and best Blast hit for CRISPR-CAS system operons.

ORF	Description based on Subsystem annotations	Accession	Identity (%)
Orf 2275	CRISPR-associated RAMP Cmr2	WP_179048547.1	92
Orf 2276	CRISPR-associated RAMP Cmr3	WP_102220820.1	95
Orf 2277	CRISPR-associated RAMP Cmr4	WP_179048549.1	94
Orf 2278	CRISPR-associated RAMP Cmr5	WP_102220822.1	95
Orf 2279	CRISPR-associated RAMP Cmr6	WP_179048551.1	78
Orf 2288	CRISPR-associated protein Cas1	WP_218653184.1	94
Orf 2289	CRISPR-associated protein Cas2	WP_179048556.1	94
Orf 3613	CRISPR-associated negative autoregulator Cas7/Cst2	WP_194144297.1	94
Orf 3614	CRISPR-associated protein Cas5	MBW4428053.1	96
Orf 5125	CRISPR-associated protein Cas2	WP_190898341.1	95
Orf 5126	CRISPR-associated protein Cas1	MBN3899475.1	93
Orf 5227	CRISPR-associated RecB family exonuclease Cas4	MBW4675378.1	91
Orf 5128	CRISPR-associated endoribonuclease Cas6	WP_096682797.1	94
Orf 5137	CRISPR-associated helicase Cas3	WP_179075640.1	97
Orf 5452	CRISPR-associated RAMP Cmr2	MBD2365142.1	92
Orf 5453	CRISPR-associated RAMP Cmr3	WP_190709980.1	95
Orf 5454	CRISPR-associated RAMP Cmr4	WP_190709978.1	93
Orf 5455	CRISPR-associated RAMP Cmr5	WP_190709976.1	95
Orf 5456	CRISPR-associated RAMP Cmr6	WP_190709975.1	91

4. Discussion

Cyanobacterial biomasses are well-documented sources of metabolites for biotechnological applications and hopefully for humankind. Some cyanobacteria species are an inherited foodstuff from ancient cultures and are still consumed nowadays [8,12,30]. In South America, edible *Nostoc* colonies, known by the vernacular name of Llayta, have been consumed since pre-Columbian times, and today they are available as dry biomass at food markets in northern Chile and southern Peru [8,17,31].

Nostoc species are morphologically diverse and ubiquitous to almost any ecosystem on Earth, including those under extreme environmental conditions. Trichomes containing vegetative and N_2-fixing heterocyst cells develop during the life cycle of diazotrophic *Nostoc* species. Filaments accumulate into a mucilaginous matrix rich in a complex polysaccharide enclosed by an outer envelope in *Nostoc* species capable of forming macrocolony. *Nostoc* macrocolonies can be observed as sheet-like or spherical forms with green to brown colorations, depending upon the species, dehydration state, and local environment [17-22,32]. In natural environments, macrocolonies of filamentous diazotrophic *Nostoc* are colonized by diverse microorganisms involved in biogeochemical cycles and ecosystem services [33]. Differences in the diversity of heterotrophic bacteria have been reported at the inner matrix, the envelope, or the surrounding environment of *Nostoc* macrocolonies. Secker et al. [20] have reported the absence of bacteria at the inner matrix of *Nostoc* macrocolonies from ephemeral wetlands in New Zealand, but high bacterial diversity at their external surface enriched with members of the genus *Sphingomonas*. Conversely, a

 doi:10.20944/preprints202202.0189.v1

different bacterial composition was found at the inner matrix, outer layer, and the littoral zone on *Nostoc* macrocolonies from Chungará Lake in northern Chile [21]. Comparatively, and using metagenomics, multiple taxonomic markers, and microscopic approaches, Satjarak et al. [22] reported high taxonomic diversity (cyanobacteria, microalgae, and anoxygenic bacterial genera) on the accompanying epimicrobiota of macroscopic dark-brown sheets of *Nostoc* from standing water pools at Parinacota (Lauca National Park, northern Chile).

An accompanying microflora has been previously observed in edible Llayta macrocolonies during attempts to isolate axenic trichomes of the dominant *Nostoc* cyanobacterium. Vilo et al. [23] have reported the identification and the draft genome of a *Bacillus* bacterium from the microbiota associated with Llayta colonies as a first approach to address physiological relationships within the Llayta microbiome and to gain insights on the survival and adaptive strategies to dryness, arsenic, and UV radiation, among other prevalent extreme environmental conditions at the Andes wetlands. The presence of *Anabaena* on the resilient microflora accompanying the filamentous *Nostoc* Llayta was coincident with previous observations reported on *Nostoc* colonies from Parinacota wetlands [22].

Metagenomics is a proper culture-independent approach to gain insights into the metabolic capabilities of Llayta cyanosphere and natural products prospection [34]. In this context, the present report provides new metagenomics-based information to improve our understanding of the Llayta cyanosphere. Previous attempts to purify axenic filaments from Llayta were unsuccessful due to the presence of high heterotrophic bacterial titers. We have applied metagenomics analyses to identify the antibiotic-resilient bacterial community on Gentamycin-treated Llayta filaments, to reconstruct genomes of the colony-forming *Nostoc* cyanobacterium and five prominent bacteria, and to detect biosynthetic gene clusters related to primary and secondary metabolism and adaptive stress strategies.

Previously, Galetovic et al. [17] reported that Llayta macrocolonies were formed by a cyanobacterium strain belonging to the *Nostoc* genus and referred to as *Nostoc* sp. Llayta. According to phylogenetic analyses based on four concatenated marker genes (*16S rRNA*, *recA*, *dnaJ*, and *gyrB*), we have confirmed this taxonomical affiliation. However, the study showed no precise placement to one species and might indicate that this Andean cyanobacterium strain would be a novel species; however, further analyses are required to confirm this hypothesis.

The genomic information on functional capabilities of the Llayta microbiome may be understood as part of the adaptive responses of this microbiome to survive and proliferate under the stressful environmental conditions found at the Andean wetlands where these macrocolonies can be found (e.g., high solar UV in-solation, desiccation, pH, osmotic and temperature changes). Andean wetlands over 4,000 m of altitude are habitats under daily and seasonal temperature variations but also under high UV irradiation [29]. Then, resilient life forms have evolved morphological and biochemical adaptive mechanisms to withstand these environmental stressors. In *Nostoc* macrocolonies, the outer envelope and the mucilaginous inner matrix are physical protecting barriers against solar radiation and loss of intracellular water. The gene mining conducted on the reconstructed genome of *Nostoc* sp. Llayta filaments demonstrated the presence of conserved lycopene cyclase *cruA*/*cruP* genes and complete *mysABCD* and *scyABCDEF* gene clusters involved in mycosporine-like amino acids and scytonemin biosynthesis, two UV-absorbing molecules located at the outer cell membrane and the intracellular compartment, respectively [35]. This genomic analysis represents the putative synthetic capabilities of *Nostoc* sp. Llayta needed for protection against photooxidative damages triggered by high-UV solar insolation prevailing at the Andes wetlands [29].

 doi:10.20944/preprints202202.0189.v1

Lasso peptides are ribosomally-synthesized and post-translationally modified small peptides with antimicrobial, antiviral, and enzymatic inhibitors [34]. The gene cluster for Lasso peptide biosynthesis includes the precursor peptide, an ATP-dependent cysteine protease, and one ATP-dependent macrolactam synthetase. The general cluster organization has been studied in Actinobacteria, Firmicutes, and Proteobacteria phyla showing differences in the presence of an ABC-transporter gene and a splitting of the B proteins in Actinobacteria in contrast to Proteobacteria cluster; however, both gene organizations can produce the Lasso peptide [36]. Gene mining on the genome of *Nostoc* sp. Llayta showed the presence of all genes needed for lasso peptide biosynthesis with a similar organization found in Actinobacteria [34,36,37]. Although the cluster in *Nostoc* sp. Llayta has a reverse direction for transcription (Fig 5 A) compared to the Actinobacteria cluster, it might produce the necessary peptides for the biosynthesis of Lasso peptide. According to phylogeny, the precursor peptide sequence is conserved in the *Nostoc* family (Fig 5. C), which might indicate the potential of the genera for the biosynthesis of Lasso peptides.

Gene mining of the assembled genome of *Nostoc* sp. Llayta showed the potential capability for capreomycidine biosynthesis and the production of members of the tuberactinomycin family of antibiotic peptides such as viomycin and capreomycin [38], but also for the biosynthesis of a microviridin precursor that would render a serine-protease inhibitor peptide and the Andean cyanobacterium as a new source of microviridins [39].

The bioinformatic analysis of the metagenome-assembled genome from *Nostoc* sp. Llayta showed the presence of CRISPR-Cas operons conferring an adaptive immunity defense against mobile genetic elements such as phages to the Andean cyanobacterium. Such defensive systems have been reported in archaea and bacteria [40,41], organized as CRISPR arrays (short, direct repeats of 20–40 bp separated by variable spacers) and Cas operons [40-42]. According to Shao et al. [43], the Cmr complex binding and cleaving of foreign RNA may involve a crRNA endoribonuclease activity coded at one of the Cmr1, Cmr3, Cmr4, or Cmr6 genes; in contrast, the Cmr2 gene is involved in binding specific targets. The crRNA transcribed by the CRISP array would bind to CAS proteins to recognize the foreign nucleic acid, which would be cleaved by Cas1 and Cas2 nucleases [44-47]. Sequences for CRISPR-CAS system operons found at the metagenome-assembled genome from *Nostoc* sp. Llayta were substantially similar sequences from CRISPR-CAS operons available at public databases.

5. Conclusions

The gentamycin-resilient microbial community associated with cyanobacterial macrocolonies of Llayta filaments was identified using a culture-independent approach. Metagenomics analyses allowed the reconstruction and draft genomes for five prominent bacteria and the genome of the macrocolony-forming cyanobacterium *Nostoc* sp. Llayta. This information provides an insight into microbial functional capabilities, biosynthetic pathways, and adaptive strategies to the environmental conditions at high-altitude Andean wetlands. Beyond the well-known, centuries-old consumption of Llayta macrocolonies, their cyanosphere opens new opportunities for biotechnological applications. The cyanosphere associated to *Nostoc* filaments of edible Llayta macrocolonies is another example of the microbial richness at the Atacama Desert deserving extensive research and environmental protection [48].

Supplementary Materials: The following document is available online at www.mdpi.com/xxx/s1, Supplementary Figure S1. Phylogenetic tree of 16S rRNA gene depicting clades of the five assembled bacteria genera, *Aquimonas*, *Blastomonas*, *Microvirga*, *Mesorhizobium*, and *Paracoccus*.

Author Contributions: Conceptualization, C.V., B.G-S.; methodology, C.V., A.G.; validation, A.G. C.V., Q.D.; formal analysis, C.V.; investigation, C.V., A.G., and B.G-S.; data curation, C.V., A.G.;

writing-original draft preparation, B.G-S., C.V.; writing-review and editing, B.G-S, C.V., A.G., Q.D.; project administration, B.G-S.; funding acquisition, B.G.-S. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by CONICYT, Chile, grant CeBiB FB-0001.

Acknowledgments: Data was generated in the Genome Sequencing Facility, which is supported by UT Health San Antonio, NIH-NCI P30 CA054174 (Cancer Center at UT Health San Antonio), NIH Shared Instrument grant 1S10OD021805-01.

Conflicts of Interest: The authors declare no conflict of interest. The funding organization had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript, or in the decision to publish the results.

References

1. Bertonio, L. Vocabulario de la Lengua Aymara. Colección Biblioteca Nacional de Chile: Impreso en la Compañía de Jesús, Perú. Electronic Document. 1612. Available online: http://www.memoriachilena.cl/602/w3-article-8656.html (accessed 11 October 2021).
2. Lagerheim, M.G. (1892). La Yuyucha. La Nuova Notarisia 3:1376-1377. https://ia601307.us.archive.org/23/items/lanotarisiacomme7293levi/lanotarisiacomme7293levi.pdf. (accessed 11 October 2021).
3. Mannheim, B. (1986). Poetic form in Guaman Poma's Wariqsa Arawi. Amerindian, 11, 41–67. https://sedyl.cnrs.fr/amerindia/articles/pdf/A_11_03.pdf. (accessed 11 October 2021).
4. Becker, E.W. (2013). "Microalgae for Human and Animal Nutrition", in Handbook of Microalgal Culture, ed. A. Richmond (Oxford: John Wiley & Sons, Ltd.), 461-503. https://doi.org/10.1002/9780470995280.ch18.
5. Gao, K. (1998). Chinese studies on the edible blue-green alga, Nostoc flagelliforme: a review. J. Appl. Phycol. 10, 37-49. https://doi.org/10.1023/A:1008014424247.
6. Qiu, B., Liu, J., Liu1, Z., Sliu. (2002). Distribution and ecology of the edible cyanobacterium Ge-Xian-Mi (Nostoc) in rice fields of Hefeng County in China. J. Appl. Phycol. 14, 423-429. https://doi.org/10.1023/A:1022198605743.
7. Gantar, M., Svirčev, Z. (2008). Microalgae and cyanobacteria: food for thought. J. Phycol. 44, 260-268. doi: 10.1111/j.1529-8817.2008.00469.x.
8. Rivera, M., Galetovic, A., Licuime, R., Gómez-Silva, B. (2018). A microethnographic and ethnobotanical approach to Llayta consumption among the Andes feeding practices. Foods 7, 202. doi:10.3390/foods7120202.
9. Vargas, M.A., Rodríguez, H., Moreno Fernández, J., Olivares, H., Del Campo, J.A., Rivas, J., Guerrero, M.G. (1998). Biochemical composition and fatty acid content of filamentous nitrogen-fixing cyanobacteria. J. Phycol. 34, 812-817. https://doi.org/10.1046/j.1529-8817.1998.340812.x.
10. Niccolai, A., Bigagli, E., Biondi, N., Rodolfi, L., Cinci, L., Luceri, C., Tredici, M.R. (2017). In vitro toxicity of microalgal and cyanobacterial strains of interest as food source. J. Appl. Phycol. 29, 199–209. https://doi.org/10.1007/s10811-016-0924-2.
11. Wells, M.L., Potin, P., Craigie, J.S., Raven, J.A., Merchant, S.S., Helliwell, K.E., Smith, A.G., Camire, M.E., Brawley, S.H. (2017). Algae as nutritional and functional food sources: revisiting our understanding. J. Appl. Phycol. 29, 949-982 DOI 10.1007/s10811-016-0974-5.
12. Li, Z., Guo, M. (2018). Oncotarget. Healthy efficacy of Nostoc commune Vaucher. https://doi.org/10.18632/oncotarget.23620.
13. Nandagopal, P., Steven, A.N., Chan, L-W., Rahmat, Z., Jamaluddin, H., Noh, N.I.M. (2021). Bioactive metabolites produced by cyanobacteria for growth adaptation and their pharmacological properties. Biology 10, 1061. https://doi.org/10.3390/biology10101061.
14. Bull, A.T., J.A. Asenjo, M. Goodfellow, Gómez-Silva, B. (2016). The Atacama Desert: Technical Resources and the Growing Importance of Novel Microbial Diversity. Ann. Rev. Microbiol. 70, 215-234. doi: 10.1146/annurev-micro-102215-095236.
15. Gómez-Silva, B. (2018). Lithobiontic life: Atacama rocks are well and alive. Antoine van Leeuwenhoek 111, 1333-1345. https://doi.org/10.1007/s10482-018-1033-9.
16. Houston, J. (2006). Variability of precipitation in the Atacama Desert: its causes and hydrological impact. Int. J. Climatol. 26, 2181-2198. DOI: 10.1002/joc.1359.
17. Galetovic, A., Araya, J.E., Gómez-Silva, B. (2017). Composición bioquímica y toxicidad de colonias comestibles de la cianobacteria andina Nostoc sp. Llayta. Rev. Chil. Nutr. 44, 360-370. http://dx.doi.org/10.4067/s0717-75182017000400360.
18. Dodds, W., Gudder, D.A., Mollenhauer, D. (1995). The ecology of *Nostoc*. J. Phycol. 31, 2-18. https://doi.org/10.1111/j.0022-3646.1995.00002.x.
19. Gao, S., Kong, Y., Yu, J. (2020). Isolation of axenic cyanobacterium and the promoting effect of associated bacterium on axenic cyanobacterium. BMC Biotechnology. 20,61. https://doi.org/10.1186/s12896-020-00656-5.
20. Secker, N.H., Chua, J.P.S., Laurie, R.E. (2017). Characterization of the Cyanobacteria and Associated Bacterial Community From an Ephemeral Wetland in New Zealand. J Phycol. 52, 761-773. doi: 10.1111/jpy.12434.
21. Aguilar, P., Dorador, C., Vila, I., Sommaruga, R. (2019). Bacterial communities associated with spherical Nostoc macrocolonies. Front. Microbiol. 10,483. doi: 10.3389/fmicb.2019.00483.

22. Satjarak, A., Graham, L.E., Piotrowski, M.J., Trest, M.T., Wilcox, L.W., Cook, M.E., Knack, J.J., Arancibia-Avila, P. (2021). Shotgun metagenomics and microscopy indicate diverse cyanophytes, other bacteria, and microeukaryotes in the epimicrobiota of a northern Chilean wetland *Nostoc* (cyanobacteria). J. Phycol. 57, 39–50. DOI: 10.1111/jpy.13084-20-040.
23. Vilo, C., Galetovic, A., Araya, J. E., Gómez-Silva, B., Dong, Q. (2015). Draft Genome Sequence of a Bacillus Bacterium from the Atacama Desert Wetlands Metagenome. Genome announcements, 3(4), e00955-15. https://doi.org/10.1128/genomeA.00955-15
24. Keegan, K. P.; Glass, E. M.; Meyer, F. MG-RAST, a metagenomics service for analysis of microbial community structure and function. In Microbial environmental genomics (MEG); Humana Press: New York, NY., USA, 2016, pp. 207-233.
25. Li, D., Luo, R., Liu, C. M., Leung, C. M., Ting, H. F., Sadakane, K., Yamashita, H., Lam, T-W. (2016). MEGAHIT v1. 0: a fast and scalable metagenome assembler driven by advanced methodologies and community practices. Methods, 102, 3-11.
26. Wattam A.R., Abraham, D., Dalay, O., Disz, T.L., Driscoll, T., Gabbard, J.L., Gillespie, J.J., Gough, R., Hix, D., Kenyon, R., Machi, D., Mao, C., Nordberg, E.K., Olson, R., Overbeek, R., Pusch, G.D., Shukla, M., Schulman, J., Stevens, R.L., Sullivan, D.E., Vonstein, V., Warren, A., Will, R., Wilson, M.J., Yoo, H.S., Zhang, C., Zhang, Y., Sobral, B.W. (2014). PATRIC, the bacterial bioinformatics database and analysis resource. Nucleic Acids Res. D581-D591. doi: 10.1093/nar/gkt1099.
27. Skinnider, M.A., Merwin, N.J., Johnston, C.W., Magarvey, N.A. (2017). PRISM 3: expanded prediction of natural product chemical structures from microbial genomes. Nucleic Acids Res. 45(W1), W49-W54. doi: 10.1093/nar/gkx320.
28. Blin, K., Shaw, S., Kloosterman, A.M., Charlop-Powers, Z., van Wezel, G.P., Medema, M.H., Weber. T. (2021). antiSMASH 6.0: improving cluster detection and comparison capabilities Nucleic Acids Res. 49(W1): W29–W35. https://doi.org/10.1093/nar/gkab335.
29. Cordero, R.R., Damiani, A., Jorquera, J., Sepúlveda, E., Caballero, M., Fernandez, S., Feron, S., Llanillo, P.J., Carrasco, J., Laroze, D., Labbe, F. (2018). Ultraviolet radiation in the Atacama Desert. Antonie Van Leeuwenhoek. 111(8), 1301-1313. doi: 10.1007/s10482-018-1075-z.
30. Aaronson, S. (1986). A role for algae as human food in antiquity. Food Foodways. 1, 311-315. https://doi.org/10.1080/07409710.1986.9961891.
31. Gómez-Silva. B., Mendizabal, I., Tapia, I., Olivares, H. (1994). Microalgas del norte de Chile. IV. Composición química de Nostoc commune Llaita. Rev. Invest. Cient. Tecnol. Cs. Mar 3, 19-25. http://bioquimica.uantof.cl/files/Publicaciones/1994%20Rev%20Invest%20%20Nostoc%20Llaita25.pdf. (accessed 11 October 2021).
32. Sand-Jensen, K. 2014. Ecophysiology of gelatinous Nostoc colonies: unprecedented slow growth and survival in resource-poor and harsh environments. Ann Bot. 114(1), 17–33. doi: 10.1093/aob/mcu085.
33. Ducklow H. (2008). Microbial services: challenges for microbial ecologists in a changing world. Aquat. Microb. Ecol. 53, 13-19. doi: 10.3354/ame01220.
34. Tietz, J.I., Schwalen, C.J., Patel, P.S., Maxson, T., Blair, P.M., Tai, H.C., Zakai, U.I., Mitchell, D.A. (2017). A new genome-mining tool redefines the lasso peptide biosynthetic landscape. Nat Chem Biol. 13(5), 470-478. doi: 10.1038/nchembio.2319.
35. Garcia-Pichel, F., Castenholz, R.W. (1993). Occurrence of UV-Absorbing, Mycosporine-Like Compounds among Cyanobacterial Isolates and an Estimate of Their Screening Capacity. Appl. Environ. Microbiol. 59, 163-169.
36. Martin-Gómez, H., Tulla-Puche, J. (2018). Lasso peptides: chemical approaches and structural elucidation. Org. Biomol. Chem. 16(28), 5065-5080. DOI https://doi.org/10.1039/C8OB01304G.
37. Mevaere, J., Goulard, C., Schneider, O., Sekurova, O.N., Ma, H., Zirah, S., Afonso, C., Rebuffat, S., Zotchev, S.B., Li, Y. (2018). An orthogonal system for heterologous expression of actinobacterial lasso peptides in Streptomyces hosts. Sci Rep. 29;8(1), 8232. doi: 10.1038/s41598-018-26620-0.
38. Fei, X., Yin, X., Zhang, L. Zabriskie, T.M. (2007). Roles of VioG and VioQ in the Incorporation and Modification of the Capreomycidine Residue in the Peptide Antibiotic Viomycin. J. Nat. Prod. 70(4), 618–622. doi:10.1021/np060605u.
39. Ziemert, N., Ishida, K., Weiz, A., Hertweck, C., Dittmann, E. (2010). Exploiting the Natural Diversity of Microviridin Gene Clusters for Discovery of Novel Tricyclic Depsipeptides. Appl. Environ. Microbiol. 76, 3568-3574. doi:10.1128/AEM.02858-09.
40. Sorek, R., Lawrence, C. M., and Wiedenheft, B. (2013). CRISPR-mediated adaptive immune systems in bacteria and archaea. Annu. Rev. Biochem. 82, 237–266. doi: 10.1146/annurev-biochem-072911-172315.
41. Grissa, I., Vergnaud, G., Pourcel, C. (2007). The CRISPRdb database and tools to display CRISPRs and to generate dictionaries of spacers and repeats. BMC Bioinformatics 8:172. doi: 10.1186/1471-2105-8-172
42. Shmakov, S. A., Sitnik, V., Makarova, K. S., Wolf, Y. I., Severinov, K. V., Koonin, E. V. (2017). The CRISPR spacer space is dominated by sequences from species-specific mobilomes. mBio 8:e01397-17. doi: 10.1128/mBio.01397-17
43. Shao, Y., Cocozaki, A. I., Ramia, N. F., Terns, R. M., Terns, M. P., Li, H. (2013). Structure of the Cmr2-Cmr3 subcomplex of the Cmr RNA silencing complex. Structure, 21(3), 376-384.
44. Garneau, J. E., Dupuis, M.-V., Villion, M., Romero, D. A., Barrangou, R., Boyaval, P., Fremaux, C., Horvath, P., Magadán, A.H., Moineau, S. (2010). The CRISPR/Cas bacterial immune system cleaves bacteriophage and plasmid DNA. Nature 468, 67–71. doi: 10.1038/nature09523
45. Takeuchi, N., Wolf, Y. I., Makarova, K. S., and Koonin, E. V. (2012). Nature and intensity of selection pressure on CRISPR-associated genes. J. Bacteriol. 194, 1216–1225. doi: 10.1128/JB.06521-11
46. Marraffini, L. A. (2015). CRISPR-Cas immunity in prokaryotes. Nature 526, 55–61. doi: 10.1038/nature15386
47. Jungblut, A.D., Raymond, F., Dion, S. Moineau , V. Mohit, G. Quang Nguyen, M. Déraspe, É. Francovic-Fontaine, C. Lovejoy, A.I. Culley M.B., Corbeil, J., Vincent, W.F. (2021). Genomic diversity and CRISPR-Cas systems in the cyanobacterium Nostoc in the High Arctic. Environ. Microbiol. 23, 2955-2968. doi: 10.1111/1462-2920.15481.

48. Gómez-Silva, B., Batista-García, R.A. (2022). The Atacama Desert: A Biodiversity Hotspot and Not Just a Mineral-Rich Region. Front. Microbiol. 13:812842. https://doi.org/10.3389/fmicb.2022.812842.