



UNIVERSIDADE FEDERAL DO MARANHÃO

Fundação Instituída nos termos da Lei nº 5.152, de 21/10/1966 – São Luís - Maranhão.

PROGRAMA DE PÓS-GRADUAÇÃO EM SAÚDE E AMBIENTE



IOLANDA KAROLINE BARROS DOS SANTOS ROCHA

**CARACTERIZAÇÃO GENÔMICA E PROSPECÇÃO *IN SILICO* DE
METABÓLITOS SECUNDÁRIOS DA CIANOBACTÉRIA *NOSTOC*
G BBB01 DO PARQUE NACIONAL DA CHAPADA DAS MESAS-MA**

SÃO LUÍS, MA

2021



UNIVERSIDADE FEDERAL DO MARANHÃO

Fundação Instituída nos termos da Lei nº 5.152, de 21/10/1966 – São Luís - Maranhão.

PROGRAMA DE PÓS-GRADUAÇÃO EM SAÚDE E AMBIENTE



IOLANDA KAROLINE BARROS DOS SANTOS ROCHA

CARACTERIZAÇÃO GENÔMICA E PROSPECÇÃO *IN SILICO* DE METABÓLITOS SECUNDÁRIOS DA CIANOBACTÉRIA *NOSTOC* GBBB01 DO PARQUE NACIONAL DA CHAPADA DAS MESAS-MA

Dissertação apresentada à Banca Examinadora de Defesa do Programa de Pós-Graduação em SAÚDE E AMBIENTE / UFMA como requisito para a obtenção do título de Mestre em Saúde e Ambiente

Orientador: Dr. Leonardo Teixeira Dall’Agnol

Coorientadora: Dr^a. Hivana Patricia Melo Barbosa Dall’Agnol

Linha de pesquisa: Determinantes ambientais e Saúde

SÃO LUÍS – MA

2021



UNIVERSIDADE FEDERAL DO MARANHÃO

Fundação Instituída nos termos da Lei nº 5.152, de 21/10/1966 – São Luís - Maranhão.

PROGRAMA DE PÓS-GRADUAÇÃO EM SAÚDE E AMBIENTE



Ficha gerada por meio do SIGAA/Biblioteca com dados fornecidos pelo(a) autor(a). Diretoria Integrada de Bibliotecas/UFMA

BARROS DOS SANTOS ROCHA, IOLANDA KAROLINE.

CARACTERIZAÇÃO GENÔMICA E PROSPECÇÃO *IN SILICO* DE METABÓLITOS SECUNDÁRIOS DA CIANOBACTÉRIA *NOSTOC* GBBB01 DO PARQUE NACIONAL DA CHAPADA DAS MESAS-MA / IOLANDA KAROLINE BARROS DOS SANTOS ROCHA. - 2021.

69 p.

Coorientador(a): Hivana Patricia Melo Barbosa Dallagnol.

Orientador(a): Leonardo Teixeira Dallagnol.

Dissertação (Mestrado) - Programa de Pós-graduação em Saúde e Ambiente/ccbs, Universidade Federal do Maranhão, UFMA, 2021.

1. Metabolitos secundários. 2. Mineração genômica. 3. Nostoc. I. Melo Barbosa Dallagnol, Hivana Patricia. II. Teixeira Dallagnol, Leonardo. III. Título.

IOLANDA KAROLINE BARROS DOS SANTOS ROCHA

CARACTERIZAÇÃO GENÔMICA E PROSPECÇÃO *IN SILICO* DE METABÓLITOS SECUNDÁRIOS DA CIANOACTÉRIA *NOSTOC* GBBB01 DO PARQUE NACIONAL DA CHAPADA DAS MESAS-MA

Dissertação apresentada à Banca Examinadora de Defesa do Programa de Pós-Graduação em SAÚDE E AMBIENTE / UFMA como requisito para a obtenção do título de Mestre em Saúde e Ambiente

Aprovada em 18 / 06 / 2021

BANCA EXAMINADORA

Prof. Dr. Leonardo Teixeira Dall’Agnol (Orientador)
Universidade Federal do Maranhão

Prof^a. Dr^a. Hivana Patricia Melo Barbosa Dall’Agnol (Coorientadora)
Universidade Federal do Maranhão

Prof^a. Dr^a. Mayara Ingrid Sousa Lima
Universidade Federal do Maranhão

Dr. Alex Ranieri Jerônimo de Lima
Instituto Butantan

Dr. Andrei Santos Siqueira
Universidade da Amazônia

*Dedico este trabalho a Deus e aos meus filhos
Maria Helena Barros dos Santos e
José Carlos Barros, a razão da minha vida.*

AGRADECIMENTOS

Agradeço primeiramente a Deus, por me resgatar todos os dias e me permitir em tempos tão difíceis concluir mais uma etapa da minha vida. A Ele dedico toda minha vida.

Aos meus pais, Sebastiana Barros da Silva, Maria do Rosário Barros dos Santos e Felisberto Rodrigues dos Santos (*in memorian*) pelo amor, carinho e dedicação. A minha tia-mãe Maria Lucia Barros dos Santos (*in memorian*), por ter me acolhido em seu lar.

Aos meus filhos Maria Helena e José Carlos por me ensinarem como é bom viver e recomeçar todos os dias, as minhas irmãs Eunice dos Santos Coelho e Dayana dos Santos Coelho por serem minhas companheiras em todos os momentos, obrigada pelo incentivo e confiança.

Ao meu esposo Carlos Antônio Santos e Santos pelo incentivo. A Cristiane de Kacia por segurar minha mão, apoiando e incentivando, e em muitos momentos deixando seus compromissos e ficando os meus filhos, para que eu pudesse finalizar este trabalho, meus singelos agradecimentos.

Aos meus maiores incentivadores e apoiadores Ana Butarelli e Lucas Salomão, vocês são inspirações exemplo a ser seguido, obrigada pelas palavras e por me resgatarem no momento da minha vida em que tudo parecia um caos, sou imensamente grata.

Ao meu orientador Leonardo Teixeira Dall’Agnol por ter aceitado me orientar, e sempre com muito carinho buscou as melhores soluções para que pudéssemos finalizar este trabalho, mesmo diante de um momento tão difícil que estamos vivendo e a professora Hivana Patricia Melo Barbosa Dall’Agnol por sua coorientação. Sou imensamente grata aos dois pela paciência e compreensão.

Agradeço em especial as minhas amigas de laboratório e de vida Mariene Oliveira e Malena Costa, por compartilharmos nossas alegrias e tristeza e em meio a tantas dificuldades dávamos boas gargalhadas, obrigada meninas vocês foram fundamentais nessa jornada e a todos os integrantes do grupo de pesquisa GB³.

Agradeço a UFMA e a CAPES pelo auxílio concedido para realização deste trabalho.

Enfim, agradeço a todos que direta ou indiretamente estiveram ao meu lado durante toda essa etapa da minha vida e caso tenha esquecido de mencionar o nome de alguém que fique registrado o meu muito obrigada a todos (as)!

*“Por vezes sentimos que aquilo que fazemos
não é senão uma gota de água no mar. Mas o
mar seria menor se lhe faltasse uma gota”.*

Madre Teresa de Calcutá

RESUMO

As cianobactérias são microrganismos procariotos fotossintéticos que são conhecidos por produzirem uma variedade de metabólitos secundários com diversas bioatividades. Apesar do estado do Maranhão apresentar uma combinação única de biodiversidade haja vista possuir três biomas brasileiros (Amazônia, Cerrado e Caatinga) e respectivas zonas de transição, poucos estudos foram realizados para caracterizar a diversidade de fotossintetizantes e seu potencial. Diante disto, este trabalho teve como objetivo caracterizar o genoma da cepa *Nostoc* sp. GBBB01 da coleção do Grupo de Pesquisa em Biodiversidade, Bioprospecção e Biotecnologia da UFMA, isolada em cultura unialgal originalmente do Parque Nacional da Chapada das Mesas, cerrado maranhense, e investigar através de mineração genômica metabólitos secundários de interesse biotecnológico. Após o sequenciamento metagenômico, foi realizada a separação e montagem dos genomas obtidos sendo então prosseguido com a etapa de anotação do genoma isolado da cepa GBBB01 utilizando a *Prokaryotic Genome Annotation Pipeline* (PGAP) e o RAST. Foi realizada a caracterização estrutural e comparativa do genoma através das ferramentas antiSMASH 5.0, NaPDoS, OrthoVven2, o MEGA X para filogenia e aplicativo Species Tree presente no KBase para filogenômica. A análise filogenética da sequência parcial de 16S rRNA de *Nostoc* sp. GBBB01 confirmou a forte relação com outras cepas dos gêneros *Nostoc* e *Desmonostoc*. Foram identificadas 25 categorias funcionais pelo servidor RAST. A busca realizada pela ferramenta antiSMASH detectou 24 agrupamentos gênicos de vias biossintéticas. Quando comparado os resultados individuais entre os genomas, os agrupamentos mais encontrados foram bacteriocinas, terpenos, glicolipídeos heterocistos, NRPS, nostoficina, microviridina k e nostopeptolídeo A2, porém somente no genoma da GBBB01 detectou as vias biossintéticas de nosperina, anabaenopeptina/nostamida A. Para os domínios C e KS foram identificados 15 e 21 respectivamente. Este é o primeiro registro de um genoma do gênero *Nostoc* do estado do Maranhão, a cepa demonstrou a presença de genes relacionados a diversas compostos com importância do ponto de vista da saúde e ambiental, além de revelar seu alto potencial em produzir compostos bioativos relevantes para aplicações biotecnológicas.

Palavras-chave: *Nostoc*, Mineração genômica, Metabólitos secundários.

ABSTRACT

Cyanobacteria are photosynthetic prokaryotic microorganisms that are known to produce a variety of secondary metabolites with diverse bioactivities. Although the state of Maranhão presents a unique combination of biodiversity since it has three Brazilian biomes (Amazon, Cerrado and Caatinga) and respective transition zones, few studies have been carried out to characterize the diversity of photosynthesizers and their potential. In view of this, this study aimed to characterize the genome of the *Nostoc* sp. GBBB01 from the collection of the UFMA Biodiversity, Bioprospecting and Biotechnology Research Group, isolated in unialgal culture originally from the Chapada das Mesas National Park, in the Cerrado of Maranhão, and investigating by means of genomic mining secondary metabolites of biotechnological interest. After metagenomic sequencing, the genomes obtained were separated and assembled and then proceeded with the step of annotating the isolated genome of the GBBB01 strain using the Prokaryotic Genome Annotation Pipeline (PGAP) and RAST. Structural and comparative characterization of the genome was performed using antiSMASH 5.0, NaPDoS, OrthoVven2, MEGA X for phylogeny and Species Tree application present in KBase for phylogenomics. Phenetic analysis of the 16S rRNA sequence of the *Nostoc* sp. GBBB01 confirmed the strong relationship with other strains of the genera *Nostoc* and *Desmonostoc*. 25 functional categories were identified by the RAST server. The search performed by the antiSMASH tool detected 24 gene clusters of biosynthetic pathways. When comparing individual results between genomes, the most common groupings were bacteriocins, terpenes, heterocyst glycolipids, NRPS, nostoficin, microviridine k and nostopeptolide A2, but only in the GBBB01 genome did we detect the biosynthetic pathways of nosperine, anabaenopeptin / nostamide A. C and KS domains were identified 15 and 21 respectively. This is the first record of a genome of the genus *Nostoc* from the state of Maranhão, the strain demonstrated the presence of genes related to several compounds with importance from the point of view of health and environment, in addition to revealing its high potential in producing bioactive compounds relevant to biotechnological applications.

Keywords: *Nostoc*, Genomic mining, Secondary metabolites.

LISTA DE FIGURAS

	Página
Figura 1	Distribuição global de cianotoxinas. 19
Figura 2	Parque Nacional da Chapada das Mesas, Maranhão, Brasil. 31
Figura 3	Árvore filogenética de <i>Nostoc</i> sp. GBBB01 utilizando o método de agrupamento de vizinhos construída no programa MEGAX, baseada na sequência parcial do 16S rRNA (alinhamento de 917pb), utilizando o gênero <i>Gloeobacter</i> como grupo externo. Sequências das outras cepas foram obtidas do GenBank são representadas com seus respectivos números. ◀ indica a linhagem <i>Nostoc</i> sp. GBBB01 35
Figura 4	Árvore filogenômica da <i>Nostoc</i> sp. GBBB01. Os genomas marcados em amarelo foram utilizados nas análises de hibridização de DNA e demais comparações. 36
Figura 5	Categorias funcionais identificadas no genoma da <i>Nostoc</i> sp. GBBB01 pelo servidor RAST e o número total de genes relacionados a cada categoria. 37
Figura 6	Diagrama de Venn ilustrando a distribuição de grupos compartilhados e específicos de genes nas cinco cepas do gênero <i>Nostoc</i> . Um total de 3.172 grupos compartilhados de grupos ortólogos foram identificados nessas cinco cepas do gênero, que foram gerados pelo servidor da Web OrthoVenn2. 38
Figura 7	Distribuição dos agrupamentos de metabólitos secundários identificados pela ferramenta antiSMASH nos cinco genomas. 40
Figura 8	Agrupamento de Nostoficina comparado com outros agrupamentos de metabólicos similares realizado pelo KnowClusterBlast do. antiSMASH. 41
Figura 9	Agrupamento de Microviridina K comparado com outros agrupamentos de metabólicos similares realizado pelo KnowClusterBlast do. antiSMASH 42
Figura 10	A agrupamento de Nostopeptolídeo A2 comparado com outros agrupamentos de metabólicos similares realizado pelo KnowClusterBlast do. antiSMASH. 42
Figura 11	Agrupamento de Nosperina comparado com outros agrupamentos de metabólicos similares realizado pelo KnowClusterBlast do. antiSMASH. 43
Figura 12	Agrupamento de Anabaenopeptina NZ857/ Nostamida A comparado com outros agrupamentos de metabólicos similares realizado pelo KnowClusterBlast do. antiSMASH. 44
Figura S1	(A) Classificação NAPDOS referente às regiões de domínio C e KS entre as espécies em estudo. (A) regiões de domínio C com a porcentagem das classes identificadas no domínio C. (B) As regiões de domínio KS com a porcentagem das classes identificadas no domínio KS. 65

LISTA DE TABELAS

		Página
Tabela 1	Metabólitos secundários tóxico já descritos na literatura para o gênero <i>Nostoc</i> .	20
Tabela 2	Características genômicas das cepas do gênero <i>Nostoc</i> utilizadas para comparação.	33
Tabela 3	Características genômicas da <i>Nostoc</i> sp. GBBB01.	34
Tabela 4	Os valores de DDH <i>in silico</i> para <i>Nostoc</i> sp. GBBB01 e outros genomas de referência do gênero <i>Nostoc</i> .	37
Tabela S1	Os genomas recuperados da cultura não axênica pelo sequenciamento <i>do Illumina MiSeq</i> .	57
Tabela S2	Sequências identificadas de CRISPR/CAS no genoma da <i>Nostoc</i> sp. GBBB01.	58
Tabela S3	Análise de identidade do gene 16S rRNA da <i>Nostoc</i> sp. GBBB01 gerados pelo BLASTn do NCBI.	60
Tabela S4	Distribuição dos principais genes ortólogos entre os domínios Processo Biológico, Função Molecular e Componente celular, gerados pelo OrthoVenn2.	60
Tabela S5	Distribuição dos principais genes únicos identificados pelo OrthoVenn2 nos genomas.	61
Tabela S6	Metabólitos secundários identificados pelo antiSMASH no genoma da <i>Nostoc</i> sp. GBBB01.	
Tabela S7	Agrupamentos gênicos biossintéticos (BGC) identificados em <i>Nostoc</i> sp. GBBB01 pelo servidor NaPDos para domínio C	66
Tabela S8	Agrupamentos gênicos biossintéticos (BGC) identificados em <i>Nostoc</i> sp. GBBB01 pelo servidor NaPDos para domínio KS	68

LISTA DE SIGLAS E ABREVIATURAS

antiSMASH	Antibiotics & Secondary Metabolite Analysis Shell
BLAST	Basic Local Alignment Search Tool
BGCs	Biosynthetic Gene Clusters
Cluster	Agrupamento
GBBB01	Linhagem do Grupo de Pesquisa em Biodiversidade, Bioprospecção e Biotecnologia 01 (<i>Nostoc</i> sp.)
GenBank	Banco de Dados de Nucleotídeos
GO	Gene Ontology
DNA	Ácido Desoxirribonucleico
NaPDoS	Natural Product Domain Seeker
NRPS	Peptídeos Sintetases Não-Ribossômicos
NCBI	National Center for Biotechnology Information
PKS	Policetídeos Sintases
PARNACM	Parque Nacional da Chapada das Mesas
RNA	Ácido ribonucleico
rRNA	RNA ribossomal
RiPP	Peptídeos sintetizados ribossomicamente e modificados pós-tradução
kb	Quilobase
MCs	Microcistinas
NODs	Nodularinas
CYNs	Cilindropermopsinas
STXs	Saxitoxinas
ATXs	Anatoxinas
BMAA	Beta- N -Metilamino-L-Alanina

SUMÁRIO

CAPÍTULO I	14
1 INTRODUÇÃO GERAL	15
1.1 Cianobactérias	15
1.2 O gênero <i>Nostoc</i>	16
1.3 Metabolismo secundário de cianobactérias	18
1.4 Vias de síntese de moléculas bioativas	21
1.5 Genômica de cianobactérias	22
1.6 Área de Estudo	24
OBJETIVOS	26
2.1 Geral	26
2.2 Específicos	26
CAPÍTULO II	27
1 INTRODUÇÃO	29
2 MATERIAIS E MÉTODOS	30
2.1 Coleta da amostra, cultivo não axênico da <i>Nostoc</i> sp. GBBB01	30
2.2 Extração de DNA, sequenciamento de genoma, montagem e binning	31
2.3 Filogenia	32
2.4 Hibridização <i>in silico</i> DNA-DNA	32
2.5 Anotação do genoma, mineração e análise comparativa	32
3 Resultado e Discussão	33
3.1 Características genômicas da <i>Nostoc</i> sp. GBBB01	33
3.2 Filogenia	34
3.4 Hibridização <i>in silico</i> DNA-DNA	36
3.5 Anotação de elementos funcionais (RAST)	37
3.6 Genômica Comparativa	38
REFERÊNCIAS	46
MATERIAL SUPLEMENTA	57

CAPÍTULO I

INTRODUÇÃO GERAL

1 INTRODUÇÃO GERAL

1.1 Cianobactérias

As cianobactérias são um grupo filogeneticamente antigo de microrganismos procarióticos fotossintéticos. Apesar de serem procarióticas, suas células apresentam um aparato fotossintético semelhante ao das plantas, haja vista terem originado os cloroplastos das algas verdes, que inclui clorofila-a e ambos os fotossistemas I e II (KOMÁREK, 2016; WHITTON, 2012).

Embora a maioria das cianobactérias realize fotossíntese oxigênica, algumas espécies desses procariontes têm a capacidade de obter energia a partir do consumo de substratos orgânicos na ausência de luz. Assim, podem sobreviver de um modo heterotrófico em locais com pouca ou nenhuma luminosidade (BERG et al 2009; YU et al., 2009).

A existência desses microrganismos na biosfera desde o início do período Pré-cambriano, e sua ampla diversidade morfológica e fisiológica, indica suas estratégias de vida biologicamente e ecologicamente bem-sucedidas, com rápidas adaptações a várias condições ambientais, em diferentes ecossistemas. Além disso, são capazes de formar relações simbióticas com fungos, algas verdes, diatomáceas, entre outros organismos dos reinos Protozoa e Plantae (KOMÁREK, 2016).

Já foram descritos mais 270 gêneros de cianobactérias com mais de 3.000 espécies em diversos habitats. Entre os ambientes onde são encontrados, estão ambientes de água doce, oceano, solo e rocha. Algumas espécies podem ser bem-sucedidas em locais altamente salinos, e toleram ambientes extremos de alta e baixa temperatura (HUANG e ZIMBA, 2019; WHITTON e POTTS, 2012; KVÍDEROVÁ et al., 2019; WHITTON e POTTS 2000).

Produzem oxigênio como subproduto do seu processo fotossintético e alguns desses microrganismos possuem capacidade de fixar nitrogênio. Dessa forma, desempenham papéis fundamentais nos ciclos biogeoquímicos do nitrogênio, carbono e oxigênio, além da produção primários dos ambientes aquáticos (BERMAN-FRANK et al., 2003; HESS, 2011). Entre os gêneros fixadoras de nitrogênio, se encontram, *Trichodesmium* sp. que é responsável por 42% do total da fixação global de nitrogênio, além de *Anabaena*, *Aulosira*, *Calothrix*, *Plectonema* e *Nostoc* (LATYSHEVA et al., 2012; BERMAN-FRANK et al., 2003; ZAHRA et al, 2020).

Com base em sua morfologia, esses microrganismos podem ser encontrados nas formas unicelulares, coloniais ou filamentosas, com uma variedade de tamanhos celulares (KOMÁREK et al., 2014). Algumas espécies dos procariontes filamentosos apresentam células

diferenciadas como acinetos e heterocistos (WHITTON; POTTON, 2000). Os acinetos funcionam como esporos de resistência, aumentam de tamanho quando em condições ambientais desfavoráveis, e são capazes de armazenar grande quantidade de reserva de nutrientes, já os heterocistos são células especializadas na fixação de nitrogênio atmosférico (ADAMS, 2000; KAPLAN-LEVY et al., 2010).

Quanto à classificação taxonômica das cianobactérias, tem sido adotada uma abordagem polifásica, que inclui caracteres morfológicos, citológicos, bioquímicos, ecológicos e moleculares. A sua classificação está distribuída em oito ordens, sendo elas, Nostocales, Chroococciopsidales, Spirulinales, Pleurocapsales, Chroococcales, Oscillatoriales, Synechococcales e Gloeobacterales (KOMÁREK et al., 2014; KOMÁREK, 2016).

A introdução de métodos genéticos, principalmente o sequenciamento molecular, para o estudo de populações de organismos, proporcionou um progresso importante no campo taxonômico. Nos últimos anos, surgiu grande quantidade de estudos baseados no sequenciamento de numerosas cepas isoladas de cianobactérias. O resultado desse progresso foi a criação e definição de muitos novos gêneros, baseados exclusivamente em suas posições isoladas ou separadas em árvores filogenéticas (KOMÁREK, 2020).

As cianobactérias possuem grande importância econômica, além da ecológica já mencionada. Elas são responsáveis por parte significativa de biomassa encontrada na biosfera, e conhecidas por produzir uma variedade de compostos bioativos, com atividades de interesse para diferentes áreas como bioenergia, biotecnologia, farmacêutica, agricultura, nutracêutica humana e animal, entre outros (ZAHRA et al., 2020). De acordo Demay et al. (2019), os gêneros de cianobactérias que mais produzem compostos com atividades potencialmente benéficas pertencem às ordens Oscillatoriales, Nostocales, Chroococcales e Synechococcales.

1.2 O gênero *Nostoc*

Dentre a diversidade de microrganismos do filo Cianobactéria, destaca-se o gênero *Nostoc*, que pertence à ordem Nostocales, e família Nostocaceae, um grupo bastante diversificado e distribuído em todo o mundo (LAUGHINGHOUSE et al., 2019; FIDOR; KONKEL; MAZUR-MARZEC, 2019). As espécies de *Nostoc* são encontradas em ambientes terrestres, águas doces, marinhas, e condições extremas (fontes termais e geleiras) (NOWRUZI et al., 2018; THUAN et al., 2019).

Esse gênero cianobacteriano representa um dos primeiros descritos, sendo um membro proeminente deste táxon polifilético, onde mais de 300 espécies já foram descritas (BORNET

e FLAHAULT, 1888; RAJANIEMI et al. 2005; KOMÁREK, 2013; SURADKAR et al., 2017). De acordo com Singh et al. (2016), esse gênero é complexo, diverso e heterogêneo, existindo uma alta diversidade morfológica.

Esses microrganismos são caracterizados por uma fisiologia versátil, apresentando um ciclo de vida complexo. A fixação de nitrogênio nesses microrganismos está restrita a células especializadas conhecidas como heterocistos. Além disso, o ciclo de vida de *Nostoc* inclui células em repouso, acinetos e hormogônios de filamentos multicelulares móveis, também conhecidos como as unidades de dispersão (MEEKS et al., 2002; ADAMS e DUGGAN, 2008). A sua reprodução é complexa podendo ser por fragmentos dos tricomas ou pela germinação dos acinetos, podendo ocorrer em vida livre ou em simbioses com plantas e fungos (BICUDO e MENEZES, 2006; GAGUNASHVILI; ANDRÉSSON, 2018).

As colônias de *Nostoc* podem crescer em tapetes macroscópicos em formas gelatinosas, podendo apresentar cor azul-esverdeada, amarela-marrom ou marrom-escura. São geralmente esféricas no início do estágio vegetativo e posteriormente, irregular e filiforme. As células vegetativas são em forma de barril a cilíndricas, e os heterocistos são de forma esférica a oval e geralmente solitários. Já os acinetos têm aproximadamente duas vezes o tamanho da célula vegetativa e podem ser encontrados em fileiras, geralmente contendo células ovais ou elipsoidais (KOMÁREK, 2013).

A característica de *Nostoc* produzir mucilagem e formar hormogônios, geralmente usada para distinguir cepas de *Nostoc* do gênero intimamente relacionado como *Anabaena* (WATERBURY, 2006; GAGUNASHVILI; ANDRÉSSON, 2018). A forma da colônia, o formato das células terminais, a presença de bainha e vesículas, bem como o ciclo de vida, são recursos adicionais usados para a identificação de alguns gêneros (RAJANIEMI et al., 2005).

O gênero como um todo exibe vasta diversidade genética e ecológica, também possui uma história de vida complexa, por isso é essencial estudar culturas por longos períodos, comparando culturas cultivadas em laboratório com amostras de ocorrência natural. Isso torna o gênero desafiador, cuja taxonomia depende igualmente da perspicácia morfológica e da caracterização molecular (SURADKAR et al., 2017).

Ao longo do tempo, novas espécies de *Nostoc* têm sido descobertas, como no trabalho de Singh et al. (2016), através de observações fenotípicas, genéticas e filogenéticas, estudaram uma cianobactéria filamentosa (cepa NE-PS) isolada de um corpo de água doce na Índia. As observações cuidadosas dos filamentos, a presença de uma bainha distinta ao longo do comprimento do tricoma, diferenças na forma e dimensões das células vegetativas, heterócitos

e acinetos forneceram sinais morfológicos confiáveis, sendo possível definir a nova espécie do gênero *Nostoc* com o nome proposto de *Nostoc punensis*, sp. nov.

Outra espécie nova descoberta se encontra no trabalho de Suradkar et al. (2017), onde estes isolaram uma cepa 9C-PST de cianobactéria, e realizaram caracterização morfológica e molecular, juntamente com uma avaliação completa da ecologia. O estilo de orientação do filamento, tipo e natureza da bainha como a distribuição e visibilidade através do tricoma e as dimensões e forma das células vegetativas e heterócitas foram avaliados por mais de um ano usando a cultura cultivada em laboratório e as amostras de ocorrência natural. As observações fenotípicas, ecológicas, genéticas e filogenéticas indicaram que a cepa 9C-PST representa uma nova espécie do gênero *Nostoc* com o nome proposto de *Nostoc thermotolerans* sp. nov.

Devido a capacidade de fixação de carbono atmosférico e nitrogênio, os microrganismos desse gênero são fundamentais em diversas comunidades microbianas, como solos, onde podem regular o pH, aumentar a concentração de carbono e nitrogênio, reduzir a quantidade de sódio em ambientes de alta salinidade, atuar como antagonistas de fitopatógenos (insetos e nematóides), degradar agrotóxicos, além de reduzir a contaminação com metais pesados como cobre e cromo (SINGH; PANDEY; SINGH, 2011 ; FIDOR; KONKEL; MAZUR-MARZEC, 2019).

1.3 Metabolismo secundário de cianobactérias

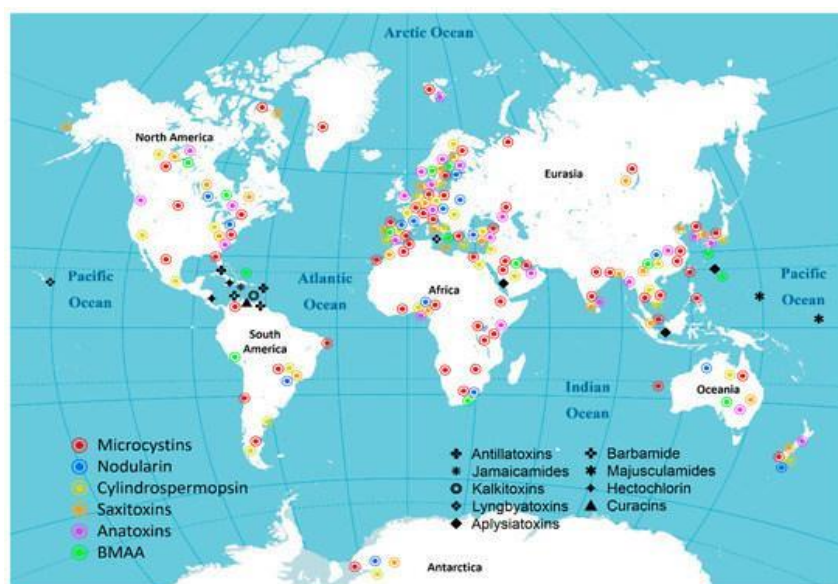
O metabolismo secundário é responsável por uma variedade de moléculas que são auxiliares aos processos essenciais para a vida, e confere vantagem ao organismo produtor. Os processos metabólicos primários são essenciais, logo são encontrados em todas as espécies. Entretanto, as vias que levam aos metabólitos secundários são especializadas, com vias particulares encontradas em um grupo taxonomicamente restrito de organismos (MASCHEK e BAKER, 2008).

Entre esses microrganismos, se encontram as cianobactérias que representam uma importante fonte de metabólitos secundários com enorme variação estrutural, possuindo classes estruturais que incluem alcalóides, depsipeptídeos, lipopeptídeos, macrolídeos/ lactonas, peptídeos, terpenos, polissacarídeos, lipídeos, policetídeos entre outros. Devido sua variação estrutural e funcional, os metabólitos exibem uma ampla gama de atividades biológicas. Essa plasticidade metabólica tem despertado interesse e há vários estudos demonstrando atividade antioxidante, antibacteriana, anticoagulante, antifúngica, antiinflamatória, antiprotzoária, antituberculose, antiviral, antitumoral, entre outros (VIJAYAKUMAR e MENAKHA, 2015; O'CONNOR, 2015; SHAH et al., 2017; DEMAY et al., 2019). Essas atividades têm potencial

para serem exploradas em campos como indústria farmacêutica, cosmética, nutracêutica humana e animal, têxtil, energética etc. (SINGH et al., 2017).

Dentre a variedade de metabólicos que esses microrganismos produzem, estão as cianotoxinas que podem causar graves problemas ambientais e de saúde (KUMAR et al., 2019; LACHAPELLE et al., 2019). Com base em seus efeitos biológicos essas podem ser classificadas como (i) hepatotoxinas, (ii) neurotoxinas, (iii) citotoxinas e (iv) dermatotoxinas (CORBEL et al., 2014). De acordo com a revisão sistemática de Du et al. (2019) as cianotoxinas mais comuns encontradas em todo globo foram as microcistinas (MCs) nodularinas (NODs), cilindropermopsinas (CYNs), saxitoxinas (STXs), anatoxinas (ATXs), e o beta- N - metilamino-L-alanina (BMAA), conforme observado na figura 1.

Figura 1- Distribuição global de cianotoxinas.



Fonte: Du et al. (2019).

Embora a América do Sul tenha tido menos relatos de toxinas cianobacterianas em comparação com outros continentes, cinco tipos comuns foram relatados: MCs, CYNs, NODs, STXs e ATXs (DU et al., 2019). Nesta região, o Brasil é o país mais afetado, e todas as cinco toxinas foram descritas como sendo encontrado em água doce (JOKELA et al., 2017; BOOPATHI e KI et al., 2014).

As microcistinas são oligopeptídeos cíclicos que foram caracterizadas pela primeira vez a partir da cianobactéria de água doce *Microcystis aeruginosa*. Muitas outras variantes estruturais foram identificadas em pelo menos dez diferentes espécies aquáticas e gêneros de cianobactérias terrestres (SIVONEN e JONES 1999; ZURAWELL et al., 2005; CARMICHAEL, 2001). Essas toxinas são consideradas as hepatotoxinas mais importantes.

Além de *Microcystis*, elas são produzidas por diversas espécies dos gêneros *Anabaena*, *Planktothrix* *Synechocystis*, *Oscillatoria*, incluindo o gênero *Nostoc* (SANTANNA et al., 2006; GALETOVIĆ et al., 2020; NOWRUZI e PORZANI, 2020). As Nodularinas são semelhantes às microcistinas, diferindo apenas na quantidade de aminoácidos em sua composição, e apresentam atividade hepatotóxicas (SANTANNA et al., 2006).

A cilindrospermopsina possui ação citotóxica e pertence a classe dos alcalóides. Elas causam interrupções de várias vias metabólicas, sendo potente inibidor de proteínas, além de serem considerada uma das cianotoxinas mais perigosas (KHOMUTOVSKA et al., 2020; SANTANNA et al., 2006). Entre as cianotoxinas que possuem ação neurotóxica, estão a anatoxina e saxitoxina. Ambas são pertencentes a classes de alcalóides, e agem sobre o sistema nervoso (CARMICHAEL et al., 2001; CHRISTENSEN e KHAN, 2020). Outra neurotoxina é o BMAA, um aminoácido que bioacumula nas cadeias alimentares (RODGERS et al., 2018).

O gênero *Nostoc* é conhecido como produtor de dezenas de metabólitos secundários, de diversas classes químicas como fenólicos, ficocianina, aminoácidos triterpenóides, ácidos graxos poliinsaturados, sulfatos, polissacarídeos, carotenoides etc. (RIMSHA et al., 2014).

Espécies do gênero *Nostoc* podem ser fontes de produtos farmacêuticos novos e úteis que são difíceis de produzir artificialmente devido à sua complexidade estrutural. Vários compostos bioativos derivados de diferentes cepas de *Nostoc* demonstraram uma nova e variada gama de atividades biológicas e estruturas químicas de novos peptídeos como, depsipeptídeos cíclicos, peptídeos cíclicos, e lipopeptídeos. Muitos deles são considerados bons candidatos para a descoberta de medicamentos, com aplicações na indústria da saúde humana e animal (GADEMANN e PORTMANN, 2008; NOWRUZI et al., 2018). Alguns metabólitos secundários produzidos por *Nostoc* com potencial para o uso como agentes farmacêuticos se encontram na Tabela 1, abaixo.

Tabela 1. Metabólitos secundários tóxico já descritos na literatura para o gênero *Nostoc*.

Classe	Metabólito	Bioatividade	Referência
Lipopeptídeo	Microginina	Inibição da enzima conversora de angiotensina (ECA)	(Zervou et al., 2020)
Peptídeos cíclico	Nodularina	Hepatotóxico	(Nowruzi e Porzani, 2021)
Depsipeptídeo cíclico	Criptoficina,	Citotóxico; Atividade fungicida	(Weiss et al., 2017; Nowruzi et al., 2018)
Heptapeptídeo não ribossômico	Nostociclopeptídeos	Atividade antitoxina	(Jokela et al., 2010)

Hexapeptídeo cíclico	Nostoficina	Citotóxico	(Fujii et al., 1999; Nowruzzi e Porzani, 2021)
Tripeptídeo não ribossômico	Nostosina	Inibidora de tripsina	(Liu et al., 2014)
Lipopeptídeo	Nostoginina,	Citotóxico	(Nowruzzi et al., 2018)
Diterpenóide	Noscomin	Atividade antibacteriana	(Carpine e Sieber, 2021)
Alcalóide	Nostocarbolina	Inibidora de tripsina	(Becher et al., 2009)
Hidrocarboneto	Paraciclofanos	Atividade Antitumoral	(Kitagaki et al., 2018)
Peptídeos cíclicos	Microcistina,	Hepatotóxico	(Zurawell et al., 2006)
Alcalóide	Saxitoxina,	Neurotóxico	(Wiese et al., 2010)

1.4 Vias de síntese de moléculas bioativas

A capacidade que as cianobactérias possuem de sintetizar numerosas substâncias químicas através de seu metabolismo secundário dá-se por serem ricas em agrupamentos de genes biossintéticos (AGB ou BGC do inglês Biosynthetic Gene Cluster) (JEONG et al., 2020). A maioria das moléculas bioativas são produzidas através de vias biossintéticas do tipo Peptídeo Sintetase não Ribossomal (NRPS – sigla em inglês), Policetídeo Sintetase (PKS), híbrido (NRPS/PKS), ou do tipo Peptídeos Sintetizados Ribossomicamente e Modificados Pós-tradução (RiPPs na sigla em inglês) (DEMAY et al., 2019; FEWER et al., 2011).

De acordo com Shih et al. (2013) o potencial genômico das cianobactérias para produzir metabólitos secundários é alto, onde 70% das cepas estudadas por esses autores apresentaram aglomerados de genes de NRPS ou PKS em seus genomas. Segundo Fidor et al. (2019), muitos peptídeos de *Nostoc* são sintetizados pelas vias NRPS ou PKS.

As NRPS são enzimas multifuncionais particularmente exclusivas que apresentam organização modular (HWANG et al., 2020). Cada módulo é responsável pela incorporação de um único aminoácido. A ordem destes módulos segue tipicamente uma regra de colinearidade, isto é, a sucessão dos módulos corresponde à ordem dos aminoácidos no produto final (KEHR et al., 2011). Em um módulo mínimo de NRPS, contém domínios funcionais específicos que são capazes de catalisar várias atividades, como a adenilação de aminoácidos, tioesterificação e formação de ligação peptídica (SCHWARZER e MARAHIEL, 2003).

As NRPS podem aceitar cerca de 300 substratos proteínogênicos e não proteínogênicos e pode conter outros domínios, introduzindo modificações personalizadas nos substratos de

aminoácidos, em contraste com a biossíntese ribossomal de peptídeos que é limitada a 20 aminoácidos proteínogênicos (GRÜNEWALD e MARAHIEL, 2006; KEHR et al., 2011).

Assim como as NRPS, as poliketídeos sintases (PKS) também são enzimas com organização modular. Elas são categorizadas em três tipos, sendo tipo I, II e III, de acordo com sua organização e mecanismos catalíticos (YU et al., 2012). Dentre eles, o PKS modular tipo I possui uma organização hierárquica em que todo o complexo enzimático é composto por várias subunidades, cada subunidade é composta por vários módulos e um módulo é composto por vários domínios (BAYLY e YADAV, 2017). Algumas cianotoxinas como, microcistina, anatoxina-a, saxitoxina e cilindrospermopsina, são sintetizados pelas vias híbridas do tipo NRPS e PKS (HUANG e ZIMBA, 2019).

Os metabólitos secundários sintetizados por vias híbridas NRPS/PKS apresentam estruturas particulares que não são alcançáveis pela síntese de peptídeo ribossomal (WELKER e DOHREN, 2006). A engenharia de vias híbridas desse tipo pode conceder acesso a novas entidades químicas que são difíceis de produzir por síntese química artificial (BOETTGER e HERTWECK, 2013).

Os peptídeos sintetizados ribossomicamente e modificados pós-tradução (RiPP), também permitem produzir compostos químicos. Sua biossíntese segue um esquema comum em que o peptídeo líder de um peptídeo precursor orienta as modificações de um único peptídeo central. Uma lista crescente de diversas modificações pós-traducionais como, macrociclizações, heterociclização, e prenilação, permite ampla diversidade química de RiPPs, a um baixo custo genético comparável aos peptídeos não-ribossômicos (ZHANG et al., 2018). Um exemplo de metabólito sintetizado ribossomicamente e modificados pós-tradução é a microviridina (HEMSCHEIDT, 2012).

1.5 Genômica de cianobactérias

Análises de genomas provavelmente fornecem o melhor registro da história biológica de uma espécie, principalmente quando se trata de microrganismo procariotos. Eles não apenas nos permitem reconstruir suas relações filogenéticas, mas também contêm informações obtidas a partir de suas interações bióticas, devido as transferências horizontais, e ambientais contínuas ao longo do tempo (MOYA et al., 2020).

O conteúdo genômico pode ser o resultado de processos neutros ou o reflexo da adaptação às diferentes condições a que um organismo está sujeito (BARRICK et al., 2009; KOONIN, 2009; TENAILLON et al., 2016). Duas estratégias de adaptação podem ser inferidas

a partir de análises genômicas de cianobactérias. Uma, fala do amplo potencial de adaptação por meio do aumento de famílias gênicas em decorrência da expansão genômica, e outra sobre a eliminação de genes dispensáveis para adaptação a determinado nicho via mecanismos de redução genômica (LARSSON et al., 2011). Dessa forma, as pressões de seleção podem causar mudanças em fatores genéticos, como tamanho do genoma, número de genes e taxas de evolução (SIMM et al., 2015).

A disponibilidade de sequências genômicas possibilitou estudar a ampla perspectiva da adaptação ambiental e as mudanças gerais nos níveis transcricional e traducional desses organismos (PRABHA et al., 2016). Além de permitir realizar análises comparativas entre genomas a fim de coletar informações para entender a evolução, funções ecofisiológicas, genes ortólogos e genes únicos, as vias metabólicas, produtos bioativos de cada grupo (tóxicos ou não tóxicos), dentre outras (PRABHA et al., 2016; MOREIRA, 2015).

Os bancos de dados do genoma de cianobactérias ainda são insuficientes quando comparado a grande diversidade do filo. Isso se deve, principalmente, pela dificuldade em obter culturas axênicas, visto que muitos desses microrganismos vivem em associações desempenhando funções complexas em consórcio. Dessa forma o material genético dos microrganismos associados também acaba sendo sequenciado, atrapalhando o processo de montagem do genoma de interesse. Assim a metagenômica se torna necessário na montagem de genomas de cianobactérias (ALVARENGA et al., 2017).

A metagenômica é uma derivação da genômica microbiana convencional, com a principal diferença que ela não necessita da obtenção de culturas puras para sequenciamento. Portanto, a metagenômica revelar os genomas da maioria dos microrganismos que não podem ser obtidos em cultura tradicional (KUNIN et al., 2008). Análise de metagenoma separa os dados de acordo com a taxonomia e assim reconstrói os genomas dos microrganismos integrantes da comunidade. Essa separação pode ser realizada de diferentes formas, seja por métodos baseados em comparação com banco de dados construídos a partir de genomas referência, seja por algoritmos que considerem as diferenças na composição das sequências e/ou perfil diferencial de cobertura de cada genoma (DRÖGE e MCHARDY, 2012; MANDE et al., 2012; ALBERTSEN et al., 2013; ALNEBERG et al., 2014).

Com os avanços das técnicas de biologia molecular, o desenvolvimento do sequenciamento de nova geração (NGS), e a bioinformática, é muito mais rápida a aquisição de dados, permitindo a análise de conjuntos de dados biológicos maiores. A comunidade de bioinformática desenvolveu um número considerável de softwares para analisar essa quantidade cada vez maior de informações biológicas (ALVARENGA et al., 2017).

Dentre as várias ferramentas desenvolvidas para minerar, ou seja, para realizar uma busca *in silico* dos metabólitos secundários existentes nos genomas de organismos sequenciados, destaca-se a ferramenta antiSMASH (*Antibiotics & Secondary Metabolite Analysis Shell*), que é utilizada para identificar e analisar *clusters* (agrupamentos) de genes biossintéticos em sequências de genomas de bactérias, fungos e plantas (BLIN et al., 2019). Outra ferramenta de mineração genômica muito utilizada é o NaPDoS (*Natural Product Domain Seeker*), onde os resultados dessa análise podem ser usados para avaliar o potencial para a biossíntese de metabólitos secundários de PKS e NRPS em organismos ou ambientes (ZIEMERT et al., 2012).

1.6 Área de Estudo

O Parque Nacional da Chapada das Mesas (PARNACM) está localizado ao sul do estado do Maranhão entre os municípios de Carolina, Estreito e Riachão (SILVA et al., 2016). É uma Unidade de Conservação de Proteção Integral criada pelo decreto presidencial de 12 de dezembro de 2005. (LIMA et al., 2013; MARQUES et al., 2014).

Localizada em uma área de transição entre três biomas - Cerrado, Amazônia e Caatinga, portanto, considerada uma área de extrema importância para conservação da biodiversidade do bioma Cerrado (SOUZA et al., 2015; SILVA et al., 2017). Abrigando assim, grande riqueza e abundância de espécies da flora e fauna, além de inúmeras nascentes de cursos d'água que abastecem três importantes bacias hidrográficas, a do Parnaíba, a do Araguaia/Tocantins e a do São Francisco (MARQUES, 2014).

Ao longo dos anos a região tem passado por aceleradas modificações antrópicas, como as plantações de soja e eucalipto, agronegócio (modernização dos processos de produção), desmatamento de grandes áreas do Cerrado, uso intenso das reservas de água e uso de agrotóxicos altamente nocivos à biodiversidade (MARQUES et al., 2014). Os quais tem contribuído significativamente para perda da biodiversidade desse bioma.

Poucos estudos foram realizados para caracterizar a diversidade e o potencial desses microrganismos essa região. Os pioneiros foram os estudos de Sousa (2019) que fez uma caracterização polifásica de duas cepas de *Nostoc* sp. (GBBB01 e GBBB02) e uma de *pantanalinema* sp. (GBBB05) e de Oliveira (2019) que fez identificação e teste de bioatividade em duas cepas de cianobactérias (*Synechococcus* sp. (GBBB07 e *Geminocytis* GBBB08) todas isoladas do PARNACM, enriquecendo os dados sobre o potencial e a diversidade desses microrganismos para o Estado do Maranhão.

Diante do exposto e pela escassez de trabalhos sobre a genômicas de cianobactérias dessa região, o presente estudo, propõem a caracterização genômica da cianobactéria *Nostoc* sp. GBBB01 e a investigação do seu potencial biotecnológico para produção de biocompostos, contribuindo para o enriquecimento de informações acerca da diversidade do filo e de seus metabólitos secundários.

OBJETIVOS

2.1 Geral

- Caracterizar o genoma da cianobactéria *Nostoc* sp. GBBB01

2.2 Específicos

- Inferir a filogenia utilizando o gene marcador 16S rRNA;
- Realizar a montagem e anotação estrutural e funcional do genoma da cepa GBBB01;
- Comparar os metabolitos secundários da cepa GBBB01 com outros genomas selecionados do gênero *Nostoc*;

CAPÍTULO II

**ARTIGO PARA SUBMISSÃO
REVISTA HARMFUL ALGAE**

Caracterização genômica e prospecção *in silico* de metabólitos secundários da cianobactéria *Nostoc* sp. GBBB01 do parque Nacional da Chapada das Mesas-MA

Iolanda Karoline Barros dos Santos Rocha^a, Hivana Patricia Melo Barbosa Dall’Agnol^b,
Leonardo Teixeira Dall’Agnol^a

^aPrograma de Pós-Graduação em Saúde e Ambiente – UFMA.

^bPrograma de Pós-Graduação em Ciências da Saúde - UFMA.

RESUMO

O gênero *Nostoc* abriga representantes cosmopolitas, comumente encontrados nos mais variados ambientes como solos, águas doces e marinhas, em simbiose com fungos (cianolíquens) e que são conhecidos por produzirem uma variedade de metabólitos secundários com diversas bioatividades. Apesar do estado do Maranhão apresentar uma combinação única de biodiversidade haja vista possuir três biomas distintos (Amazônia, Cerrado e Caatinga) e respectivas zonas de transição, poucos estudos foram realizados para caracterizar a diversidade microbiana, principalmente de fotossintetizantes, e seu potencial. Diante disto, este trabalho teve como objetivo caracterizar o genoma da cepa *Nostoc* sp. GBBB01 mantida em cultura unialgal e original do Parque Nacional da Chapada das Mesas, cerrado maranhense, e investigar através de mineração genômica vias de agrupamentos gênicos biossintéticos de interesse biotecnológico. Após o sequenciamento metagenômico do consórcio, foi realizada a separação e montagem dos genomas obtidos através do software MetaWRAP sendo então realizada a etapa de anotação do genoma da cepa *Nostoc* GBBB01 utilizando a *Prokaryotic Genome Annotation Pipeline* (PGAP) e o RAST. Foi realizada a caracterização estrutural através das ferramentas antiSMASH 5.0, NaPDoS, OrthoVven2, CRISPRDetect, PHASTER, além do software MEGA X e Species Tree presente no KBase para filogenômica e filogenia, respectivamente. A análise filogenética da sequência parcial de 16S rRNA de *Nostoc* sp. GBBB01 confirmou a forte relação com outras cepas dos gêneros *Nostoc* e *Desmonostoc*, sendo ratificados pela análise filogenômica, onde o genoma da cepa GBBB01 agrupou em um clado com espécies do mesmo gênero (*Nostoc punctiforme* PCC 73102 e *Nostoc* sp. ATCC5389). Foram identificadas 25 categorias funcionais pelo servidor RAST e 24 agrupamentos gênicos de vias biossintéticos pelo antiSMASH. Quando comparado os resultados individuais entre os genomas de referência, os agrupamentos mais encontrados foram bacteriocinas, terpenos, glicolípídeos heterocistos, vias do tipo NRPS, nostoficina, microviridina k e nostopeptolídeo

A2, sendo exclusivas de GBBB01 as vias biossintéticas de nosperina e anabaenopeptina/nostamida A. Através do NaPDoS foram identificados 15 e 21 domínios do tipo C e KS, respectivamente. Este é o primeiro registro de um genoma do gênero *Nostoc* do estado do Maranhão, a cepa demonstrou a presença de genes relacionados a diversas compostos com importância do ponto de vista da saúde e ambiental, além de revelar seu alto potencial em produzir compostos bioativos relevantes para aplicações biotecnológicas.

Palavras-chave: *Nostoc*, Mineração genômica, Metabolitos secundários.

1 INTRODUÇÃO

O gênero cianobacteriano *Nostoc* abriga representantes cosmopolitas, comumente encontrados no solo, águas doces e marinhas, e em associações simbióticas com plantas e fungos (Nowruzi et al., 2018; Thuan et al., 2019). Esses microrganismos são fixadores de nitrogênio e desempenham papéis fundamentais no ciclo biogeoquímico do nitrogênio, carbono e oxigênio contribuindo para a evolução da vida em nosso planeta (Berman-Frank et al., 2003; Gademann; Portmann, 2008; Zahra et al, 2020). Além disso, esses procariontes filamentosos são conhecidos por produzirem diferentes metabólitos secundários, com importantes atividades de interesse biotecnológico (Dittmann et al. 2001; Nowruzi e Porzani, 2021).

Entre os metabólitos produzidos estão, terpenóides, alcalóides, ácidos graxos poliinsaturados, polissacarídeos, carotenóides entre outros (O'Connor, 2015; Demay et al., 2019; Findor et al., 2019). Além disso, existem alguns metabólitos tóxicos produzidos por *Nostoc* que podem ser úteis como agentes farmacêuticos sendo eles, microcistina, saxitoxina, cianovirina-N, criptoficina, entre outros. Esses compostos possuem ampla bioatividade como, hepatotóxica, neurotóxica, atividade antiviral, citotóxica, antibacteriana e antifúngica (Dittmann e Börner, 2005; Nowruzi e Porzani, 2021; Batsalova et al., 2016; Nowruzi et al., 2018; Weiss et al., 2017).

A maioria dos produtos naturais produzidos por cianobactérias são sintetizados por vias peptídeos sintetases não ribossomais (NRPS), policetídeo sintetase (PKS) ou híbrido (NRPS/PKS) (Welker; Döhren, 2006). Essas vias são responsáveis pela capacidade que esses microrganismos possuem de sintetizarem numerosas substâncias. O gênero *Nostoc* é um exemplo dos produtores mais prolíficos de NRPs / PKs (Findor et al.,2019), o que só reforça o quanto esse gênero é um potencial a ser revelado, pois a capacidade de produzirem esses agrupamentos é específico de cada cepa (Welker; Dohren,2006).

Para estudar essas vias responsáveis pela produção de diversos metabólitos secundários, são necessários ferramentas de bioinformática que realizam buscas no genoma, a fim de detectar os possíveis metabólitos que esses microrganismos possam produzir. Por meio dessas ferramentas é possível analisar um conjunto de dados biológicos maiores, e assim estudar vários grupos de genes que codificam as vias biossintéticas para esses metabólitos (Wilkinson e Micklefield, 2007; Alvarenga et al., 2017; Calteau, et al., 2014).

O Parque Nacional da Chapada das Mesas (PARNACM), está localizado numa área de transição entre três biomas, Cerrado, Amazônia e Caatinga, possui uma alta biodiversidade e abriga grande variedade de microrganismos cianobacterianos de interesse (Souza et al., 2015; Silva et al., 2017). Entretanto, poucos estudos foram realizados para caracterizar a diversidade e o potencial desses microrganismos para essa região.

Apesar da importância das cianobactérias como produtoras de compostos bioativos, no PARNACM muitos desses microrganismos permanecem inexplorados no que se refere ao estudo genômico. Assim, este trabalho buscou caracterizar o genoma da cepa *Nostoc* sp. GBBB01 da coleção do Grupo de Pesquisa em Biodiversidade, Bioprospecção e Biotecnologia da UFMA, isolada em cultura unialgal originalmente do Parque Nacional da Chapada das Mesas, cerrado maranhense, e investigar através de mineração genômica metabólitos secundários de interesse biotecnológico.

2 MATERIAIS E MÉTODOS

2.1 Coleta da amostra, cultivo não axênico da *Nostoc* sp. GBBB01

O isolado da linhagem *Nostoc* sp. GBBB01 foi obtida de amostra do sedimento coletada em março de 2017 (período chuvoso), na cachoeira de São Romão, localizada no Parque Nacional da Chapada das Mesas (S07°02.6575/W047°30.4508), município de Carolina, Maranhão, Brasil (Fig.2). Coletou-se sedimento superficial (20 cm superior) do leito da cachoeira com um falcon estéril de 50mL. As amostras foram inoculadas em Erlenmeyers contendo meio BG-11 numa proporção de 1:10 (Watanabe e Tanabe, 2013) e incubadas em câmara de fotoperíodo sob iluminação $3-15 \mu\text{mol.m}^{-2}.\text{s}^{-1}$, 28°C e regime de iluminação de 13/11 claro/escuro até a obtenção de biomassa bacteriana. Para isolamento de culturas unialgais, foram realizadas diluições seriadas seguidas da técnica de pescaria e finalmente as amostras selecionadas foram semeadas em meio BG-11 sólido. Todo o processo foi acompanhado por análises ao microscópio para acompanhar o progresso do isolamento e controle de contaminações.

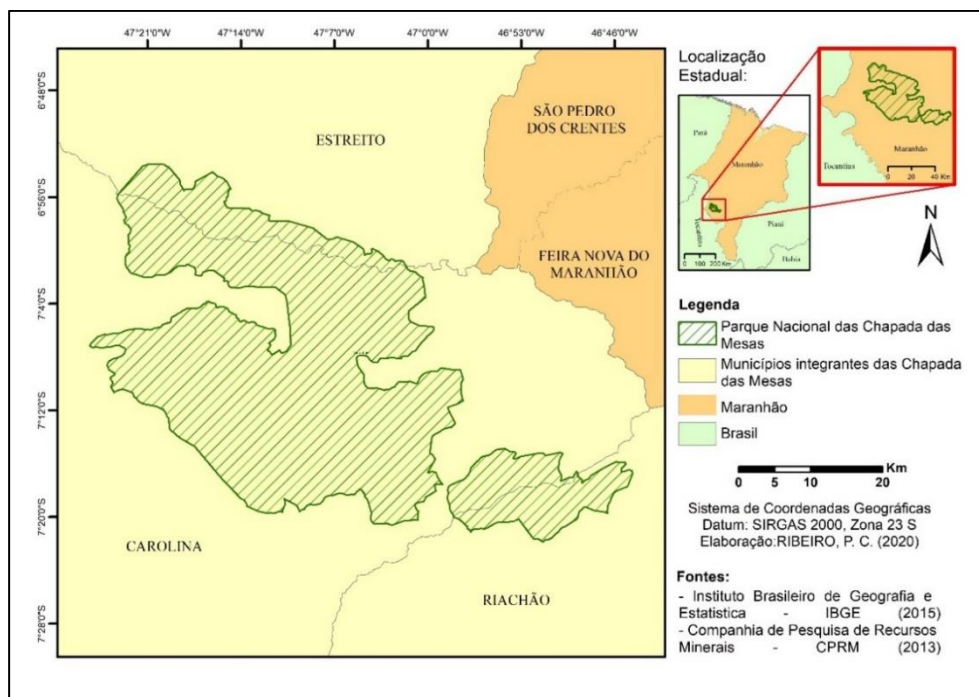


Fig. 2. Parque Nacional da Chapada das Mesas, Maranhão, Brasil.

2.2 Extração de DNA, sequenciamento de genoma, montagem e binning

O DNA foi extraído utilizando o *PowerPlant kit* (MoBio, California, USA), de acordo com as instruções do fabricante e armazenado a -20°C . A biblioteca genômica foi preparada com *Nextera DNA Sample Prep kit* (Illumina, Inc., San Diego, EUA), conforme o protocolo do fabricante. Posteriormente a quantificação e normalização, a biblioteca foi sequenciada na plataforma MiSeq/Illumina. O sequenciamento foi realizado utilizando o *MiSeq Reagent kit v2* (Illumina) de 500 ciclos com leituras *Paired-end* de 250 nucleotídeos cada. Em média as corridas geradas foram $>75\%$ de sequências com pontuação de qualidade acima de 30 (Q30).

Análise e controle de qualidade, foram realizadas através do módulo *Read_qc* do *Metawrap* (Uritskiy et al., 2018), no qual, as sequências foram trimadas e filtradas, baseadas na qualidade PHRED como parâmetro, garantindo que apenas as sequências de alta qualidade (Phred score ≥ 30) fossem conservadas. Após a obtenção das reads filtradas, os contigs foram montados através do módulo *Assembly* do *Metawrap* utilizando o *metaSPAdes v. 3.13* (Nurk et al., 2017).

O isolamento do genoma foi realizado através do módulo *Binning*, onde foram realizadas três rodadas de *binning* diferentes, uma com o *CONCOCT* (Alneberg et al., 2013), outra utilizando o *MaxBin2* (Wu et al. 2015) e outra utilizando o *MetaBAT2* (Kang et al., 2019). Os três resultados obtidos foram comparados através do módulo *Bin_refinement*, dessa forma foram selecionados os *bins* de maior qualidade. Para anotar e classificar taxonomicamente os

bins obtidos, as sequências foram submetidas ao *Prokaryotic Genome Annotation Pipeline* (PGAP) (Tatusova et al. 2016) do NCBI.

2.3 Filogenia

A sequência do gene 16S rRNA da linhagem *Nostoc* sp. GBBB01 foi comparada com 26 sequências parciais de algumas linhagens da família Merismopediaceae, Nostocaceae, Aphanizomenonaceae, Leptolyngbyaceae e o *Gleobacter* como grupo externo, obtidas do GenBank. Posteriormente foram alinhadas e editadas manualmente usando o módulo ClustalW do *software* BioEdit 7.1.3.0 (Hall, 2011). A partir do alinhamento resultante, a árvore filogenética foi construída pelo método de Neighbor-Joining (Saitou; Nei, 1987), com o modelo Kimura-2-Parametros (Kimura,1980) e 1.000 replicatas de bootstrap (Felsenstein, 1985), utilizando o *software* MEGA X versão 10.1 (Kumar et al., 2018). Árvore filogenômica foi construída usando os cinco genomas do gênero *Nostoc* comparados e outros definidos pelo Species Tree App presente no KBase (Arkin et al., 2018).

2.4 Hibridização *in silico* DNA-DNA

Para o delineamento das espécies com base no genoma, a hibridização *in silico* DNA-DNA (DDH) foi realizada usando a Calculadora de distância genoma a genoma (GGDC) (<http://ggdc.dsmz.de/home.php>). O valor de corte foi 70% usado como critério de delimitação das espécies, sendo valores acima de 30% considerados como pertencentes a mesma espécie e entre 30 e 20% do mesmo gênero (Meier-Kolthoff et al.,2013).

2.5 Anotação do genoma, mineração e análise comparativa

A identificação das classes funcionais e estruturais do genoma da *Nostoc* sp. GBBB01 foram obtidas através da anotação automática do RAST (Aziz et al., 2018; Overbeek et al., 2014). Para a comparação genômica da linhagem *Nostoc* sp. GBBB01, foram obtidos quatro genomas de referências do gênero *Nostoc*, do banco de dados do GenBank (Tabela 2). A ferramenta OrthoVenn2 (<https://orthovenn2.bioinfotoolkists.net>) versão bacteriana (XU et al., 2019) foi utilizada para comparação e análise de clusters ortólogos do genoma inteiro das linhagens.

Tabela 2. Características genômicas das cepas do gênero *Nostoc* utilizadas para comparação.

Nome do Organismo	Tamanho (Mb)	GC%	CDS	Cobertura (X)	contigs	Centro responsável
<i>Nostoc punctiforme</i> PCC 73102 (CP001037.1)*	9,06	41,34	6.989	-	6	US DOE Joint Genome Institute
<i>Nostoc</i> sp. ATCC 53789 (CP046703.1)*	8,65	41,38	6.847	38	13	Universidade TU Dortmund
<i>Nostoc</i> sp. CENA543 (CP023278.1)*	7,21	40,80	5.818	170,0	6	Universidade de Helsinque
<i>Nostoc</i> sp. NIES-2111 (AP018184.1)*	7,88	40,55	6.360	83	11	Toyohashi University of Technology

*Número de Acesso

Fonte: Centro Nacional de Informações sobre Biotecnologia.

A identificação e classificação de agrupamentos gênicos de vias foi obtida através da ferramenta AntiSMASH 5.0 versões bacteriana (Blin et al., 2019), com todas as opções selecionadas. O Servidor NaPDoS (Ziemert et al., 2012) foi utilizado para detecção de domínios C e KS. A ferramenta CRISPRCasFinder (Couvin et al., 2018) foi utilizada para prever agrupamentos CRISPR e genes CAS no genoma, e a ferramenta PHASTER (Zhou et al., 2011; Arndt et al., 2016) utilizou-se para identificar a presença de sequências referentes à profagos nos genomas. Para as demais ferramentas foram utilizadas as configurações padrões.

3 Resultado e Discussão

3.1 Características genômicas da *Nostoc* sp. GBBB01

O sequenciamento da cultura não axênica utilizando a plataforma *Illumina MiSeq* resultou em seis *bins* bacterianos relacionados aos táxons: (i) Alphaproteobacteria, (ii) *Brevundimonas*, (iii) *Nostoc*, (iv) *Leptospira*, (v) Gemmatimonadaceae e (vi) *Hydrogenophaga*. Estes apresentaram valores de completude variando de 71,1% a 100% e contaminação de 0% a 2,1% (Tabela Suplementar S1). Um total de 7.141.924 *reads* foi gerado resultando em um conjunto 500 pb com *read1+ read2*, desses seis genomas recuperados obteve-se os dados do

genoma de interesse, da cianobactéria *Nostoc*, este por sua vez resultou em um genoma com uma cobertura de 77x, com 8.131.703 pb conforme dados da Tabela 3. A busca realizada pelo PHASTER (Zhou et al., 2011; Arndt et al., 2016) não identificou nenhuma sequência de profagos no genoma bacteriano, a ferramenta CRISPRCasFinder (Couvin et al., 2018) identificou 30 sequências relacionadas ao CRISPR e nenhuma relacionada à proteína Cas no genoma da *Nostoc* sp. GBBB01 (Tabela Suplementar S2).

Tabela 3. Características genômicas da *Nostoc* sp. GBBB01.

Características	Cromossomo
Cepa	<i>Nostoc</i> sp. GBBB01
Número de Contigs	135
Valor do L50	17
Valor do N50 (pb)	154.223 pb
Completeness	99.55%
Contaminação	2.1%
Cobertura	77x
Porcentagem G + C	40,74%
Tamanho do genoma (bp)	8.131.703
N. de Protein-Coding Genes (CDS)¹	6.513
rRNAs¹	84
5S rRNAs	2
16S rRNAs	2
23S rRNA	2
tRNAs	74
ncRNAs	4
Pseudo genes	253
CRISPR^{2*}	30 sequências
CAS²	0
Fago³	0

As estatísticas do genoma foram obtidas por meio do CheckM. ¹ Pipeline de Anotação do Genoma Procariótico (PGAP); ²CRISPRCasFinder; ³PHASTER; 2 *: Tabela Suplementar S2.

3.2 Filogenia

O sequenciamento do gene 16S rRNA da cepa *Nostoc* sp. GBBB01 gerou uma sequência de 917 pb. Utilizando a ferramenta de alinhamento BLAST (Zhang et al., 2000) a sequência da cepa apresentou 99,67% identidade com a *Desmonostoc muscorum* TAU-MAC 0699, *Nostoc linckia* CALU 129 e *Desmonostoc* sp. PCC 8306. (Tabela Suplementar S3).

A árvore filogenética foi construída com 26 sequências de 16S rRNA obtidas a partir do banco de dados do NCBI. A sequência agrupou em um clado com 75% de confiança com as sequências referentes aos gêneros *Nostoc* e *Desmonostoc* (Fig. 3), demonstrando assim, que há uma forte relação filogenética entre elas. Os gêneros *Nostoc* e *Desmonostoc* são fortemente

relacionados e que, em outros trabalhos, também foram agrupados em clados com alto valor de confiança (Hrouzek et al., 2013; Saraf et al., 2018). Importante ressaltar que foram utilizadas sequências parciais do gene 16S rRNA, dessa forma, a abordagem que utiliza apenas um marcador filogenético para inferência da filogenia do grupo demonstra não ser resolutiva para este grupo.

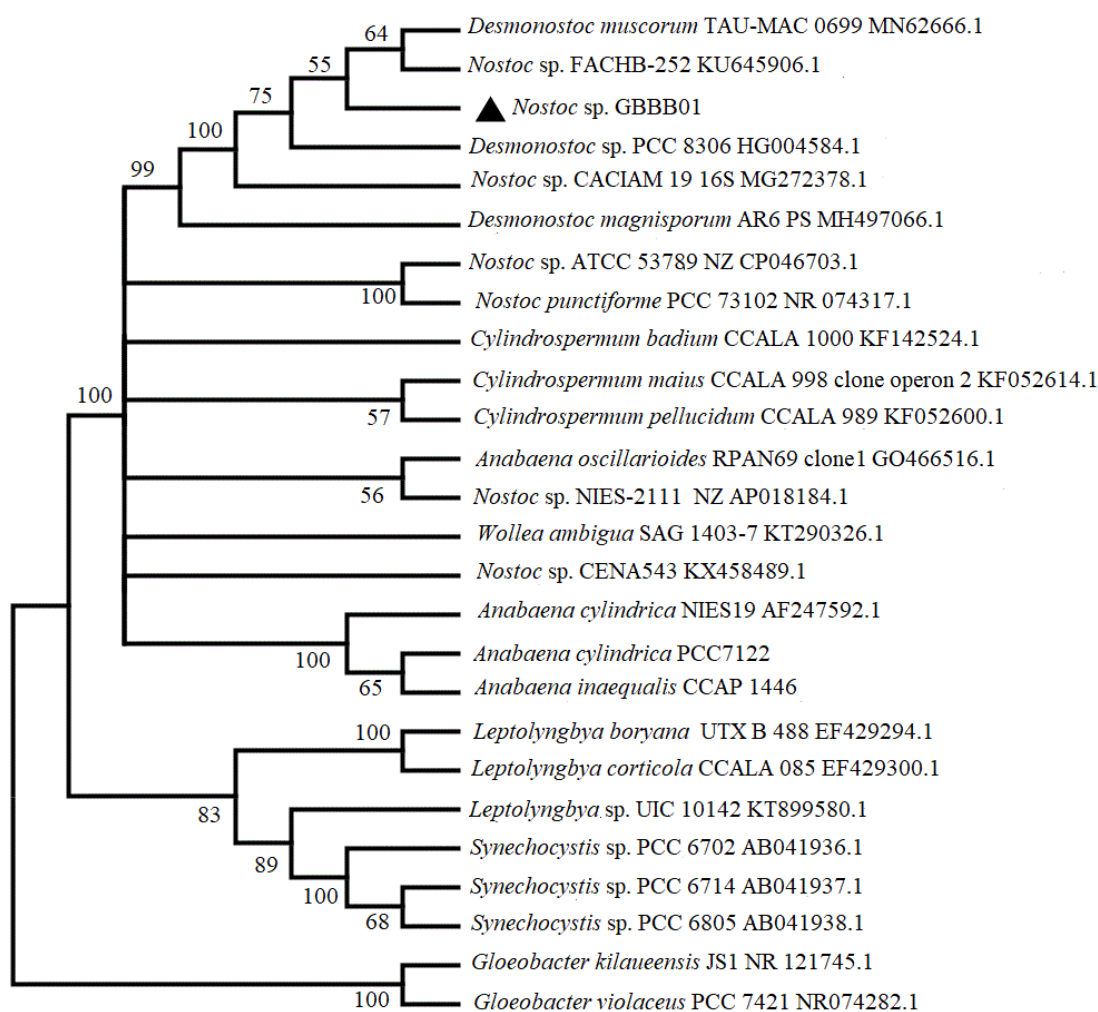


Fig.3. Árvore filogenética de *Nostoc* sp. GBBB01 utilizando o método de agrupamento de vizinhos construída no programa MEGAX, baseada na sequência parcial do 16S rRNA (alinhamento de 917pb), utilizando o gênero *Gloeobacter* como grupo externo. Sequências das outras cepas foram obtidas do GenBank são representadas com seus respectivos números. ◀ indica a linhagem *Nostoc* sp. GBBB01

Na análise filogenômica o genoma da cepa GBBB01 agrupou em um clado contendo com dois representantes do mesmo gênero (*Nostoc punctiforme* PCC 73102 e *Nostoc* sp. ATCC5389) com 100% de confiança (Fig. 4), demonstrando a forte relação filogenética entre as cepas. *Nostoc* é um gênero desafiador com características difícil de diferenciar, e sua plasticidade o torna taxonomicamente complicado (Bagchi et al., 2017). Por isso, análise

filogenômica é uma importante ferramenta para classificação filogenética deste grupo tão complexo.

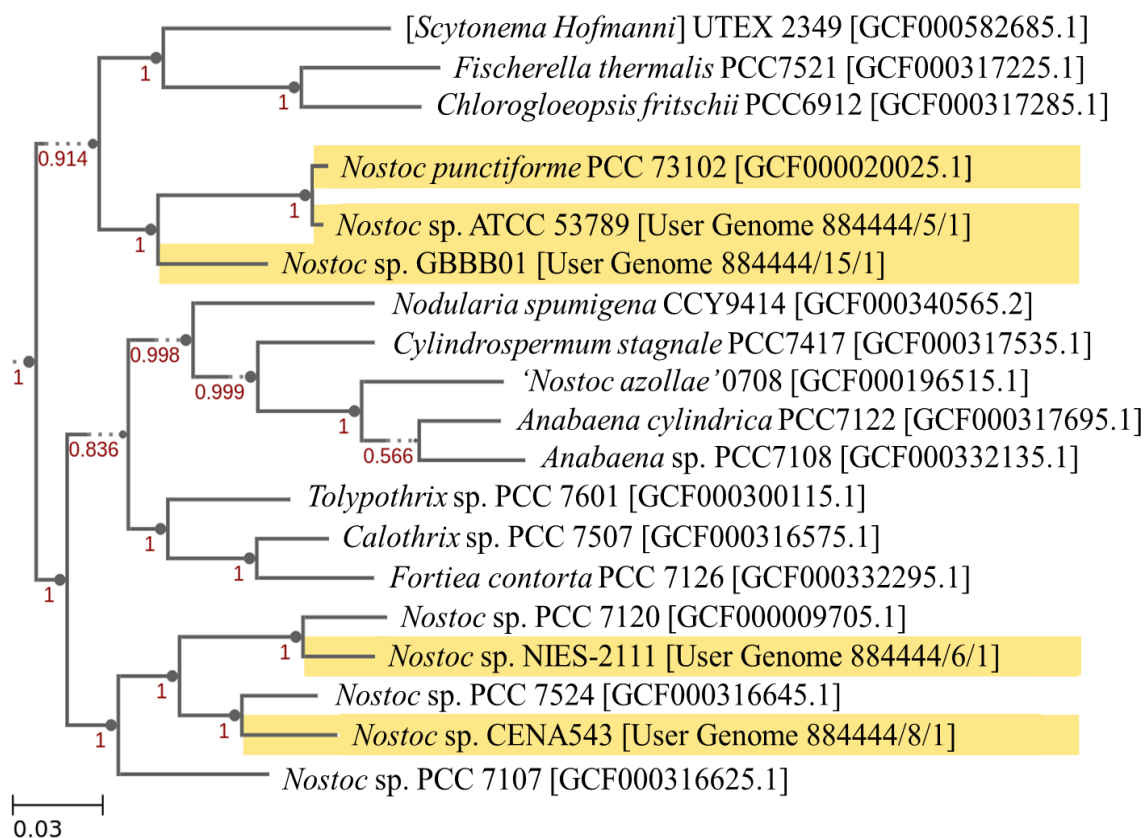


Fig.4. Árvore filogenômica da *Nostoc* sp. GBBB01. Os genomas marcados em amarelo foram utilizados nas análises de hibridização de DNA e demais comparações.

3.3 Hibridização *in silico* DNA-DNA

A análise de hibridização DNA-DNA foi utilizada para analisar a semelhança genética entre espécies do gênero *Nostoc* sp. a fim de determinar a distância genética entre as espécies. O resultado da análise revelou que valores os de hibridização em todas as espécies analisadas, foram abaixo de 22,80% (Tabela 4), sugerindo que a cepa GBBB01 não se assemelha às mesmas espécies comparadas, visto que valores abaixo de 70%, é considerado como espécies distintas (Meier-Kolthoff et al.,2013).

Tabela 4. Os valores de DDH *in silico* de *Nostoc* sp. GBBB01 e dos outros genomas de referência

Genoma de referência (Número de acesso)	Diferença em % G+C	DDH (%)	Distância	Probabilidade que DDH > 70%
<i>Nostoc punctiforme</i> PCC 73102 (CP001037)	0.68	21.80%	0.66	0
<i>Nostoc</i> sp. CENA543 (CP023278)	0.10	14.00%	0.92	0
<i>Nostoc</i> sp. NIES-2111 (AP018184)	0.11	14.20%	0.91	0
<i>Nostoc</i> sp. ATCC 53789 (CP046703)	0.69	22.80%	0.64	0

3.4 Anotação de elementos funcionais (RAST)

Anotação automática do genoma realizada pelo servidor RAST, encontrou cerca de 25 categorias funcionais no genoma de GBBB01 (Fig.5), das quais, as representativas foram: aminoácidos e derivados com (160) genes, carboidratos com (148), cofatores, vitaminas, grupos protéticos, pigmentos com (143), metabolismo de proteínas com (140). Na categoria metabólitos secundários foram identificados (13) genes, dos quais, cinco (5) estão relacionados aos hormônios vegetais (auxina), um (1) alcalóides vegetais (L-lisina), quatro (4) na síntese de microsporina (UV), 1 relacionada a biossíntese de vitamina C e um (1) de vitamina D.

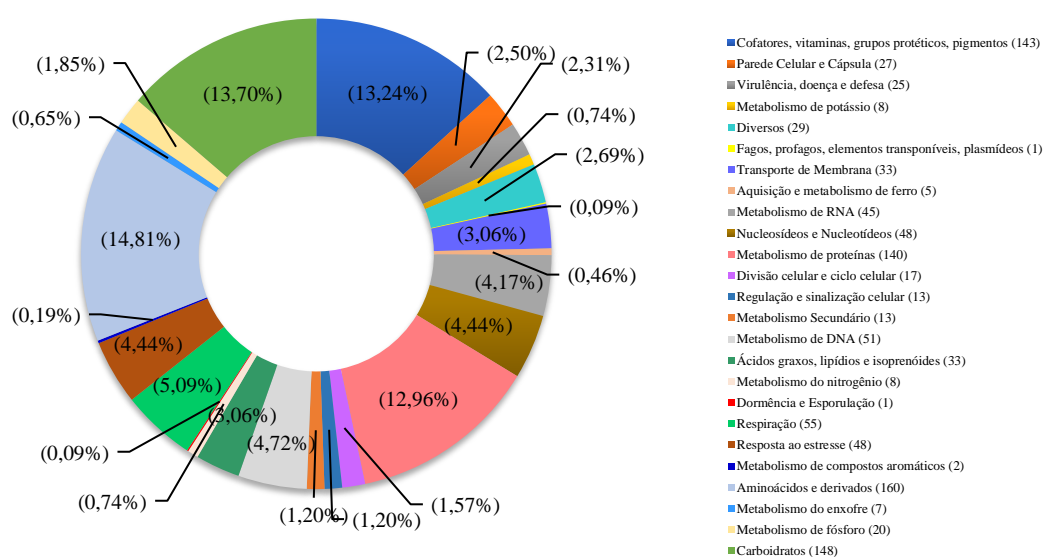


Fig.5. Categorias funcionais identificadas no genoma da *Nostoc* sp. GBBB01 pelo servidor RAST e o número total de genes relacionados a cada categoria.

3.5 Genômica Comparativa

A comparação dos genomas através da ferramenta OrthoVenn2, detectou um número total de agrupamentos ortólogos que variou de 4.549 a 5.022. A cepa *Nostoc punctiforme* PCC 73102 apresentou o maior número de genes, e a cepa *Nostoc* sp. CENA543 o menor número de genes. Um total de 3.172 grupo de genes ortólogos compartilhados foram identificados entre as cinco espécies do gênero *Nostoc*. A cepa *Nostoc* GBBB01 teve o maior número de famílias de genes únicos (n = 41), e a *Nostoc* sp. ATCC 53789 teve o menor número de genes únicos (n = 11) (Fig. 6).

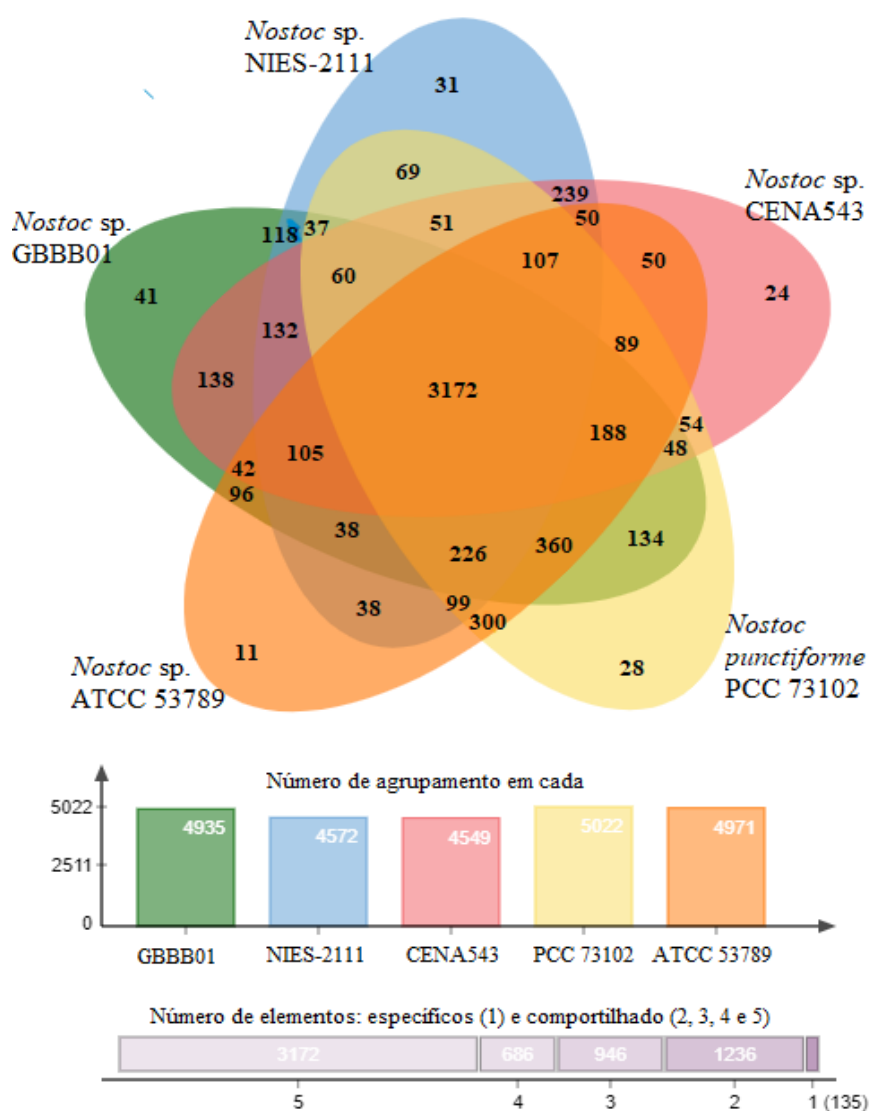


Fig. 6. Diagrama de Venn ilustrando a distribuição de grupos compartilhados e específicos de genes nas cinco cepas do gênero *Nostoc*. Um total de 3.172 grupos compartilhados de grupos ortólogos foram identificados nessas cinco cepas do gênero, que foram gerados pelo servidor da Web OrthoVenn2.

Os genes ortólogos são genes em diferentes espécies que descendem de um gene comum no mesmo ancestral. As análises realizadas, dentre os cinco genomas, revelaram que a maioria dos genes ortólogos pertencem ao processo biológico (GO:0008150) com 17,9% dos genes, além disso destaca-se 10,3% dos genes envolvidos no processo metabólico do composto de nitrogênio. Segundo Fidor (2019) as espécies do gênero *Nostoc* são conhecidas por serem fixadoras de nitrogênio, necessitando assim de número considerado de genes para realizem esse processo, que é de grande relevância para esses microrganismos e para natureza. Em relação ao domínio função molecular, a atividade de hidrolase (GO:0016787) representou 17,6% dos genes referentes à esse processo (Tabela Suplementar S4). A maior parte dos genes únicos observados nos genomas estavam envolvidos no processo biológico e metabólico. Enquanto no domínio função molecular a maioria estavam relacionados com atividade de transferase (Tabela Suplementar S5).

A análise realizada pela ferramenta antiSMASH 5.0, resultou na identificação de 24 grupos de genes biossintéticos para o genoma da GBBB01, que incluíram: seis agrupamentos possivelmente para produção de bacteriocinas; quatro agrupamentos de terpeno; três módulo de peptídeo sintetase não ribossômico (NRPS); um agrupamento (hglE-KS /T1PKS) e outro (hglE-KS) ambos possivelmente envolvidos na produção de glicolipídeos heterocistos; um agrupamento misto (transAT-PKS, T3PKS, PKS, T1PKS e NRPS) possivelmente envolvido na biossíntese de nosperina; um agrupamento para produção de nostoficina; um agrupamento para biossíntese de tiacumicina B; um agrupamento para produção de tricamida; um agrupamento (NRPS/T1PKS) para produção de nostopeptolideo A2; um agrupamento contendo via biossintética de NRPS que podem ser responsáveis pela produção de anabaenopectina NZ857 e nostamida A, e um para biossíntese de microviriduna k (Tabela Suplementar S6).

Quando comparado os resultados individuais dos genomas, os agrupamentos mais encontrados foram bacteriocinas, terpenos, glicolipídeos heterocistos, NRPS, nostoficina, microviridina k e nostopeptolídeo A2. Dentre os agrupamentos gênicos observados somente no genoma da GBBB01 destaca-se um possivelmente relacionado à síntese de nosperina, outro relacionado à síntese anabaenopectina NZ857 e nostamida A e uma relacionada a síntese de fosfato (Fig. 7).

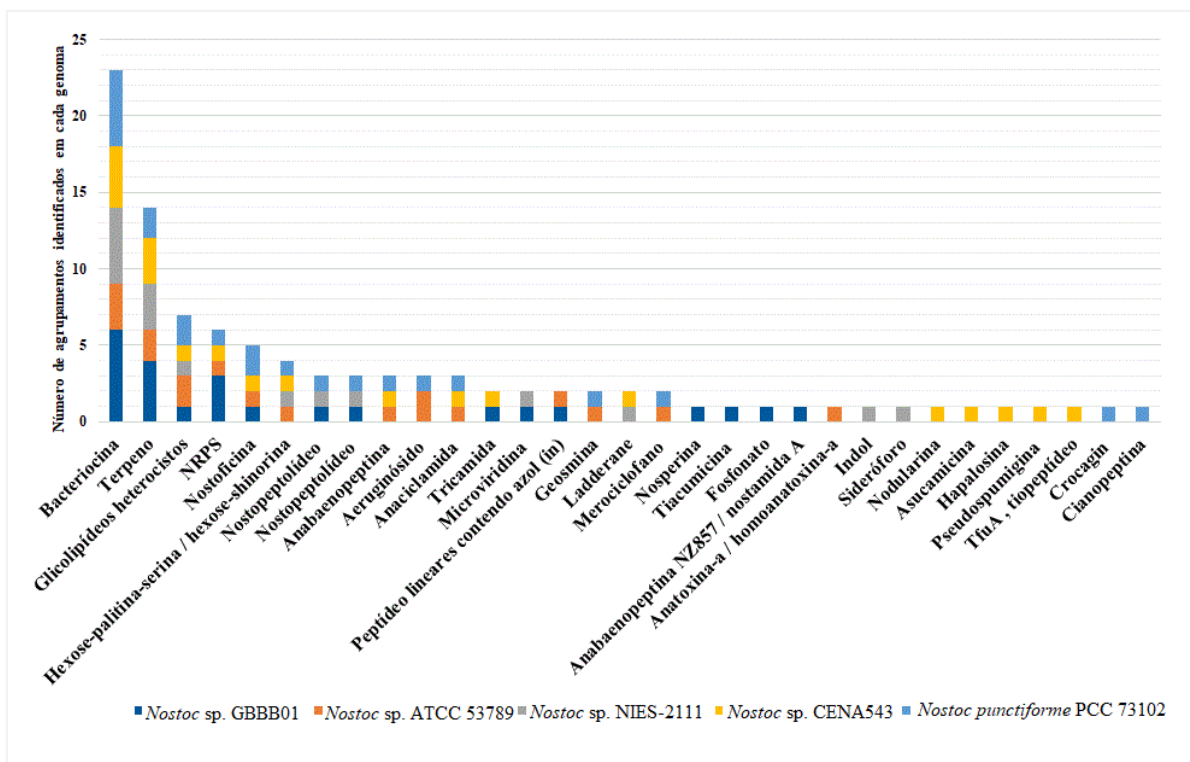


Fig.7. Distribuição dos agrupamentos de metabólitos secundários identificados pela ferramenta antiSMASH nos cinco genomas.

As bacteriocinas são um grupo diverso de peptídeos antimicrobianos sintetizados ribossomicamente, produzidos a partir de uma variedade de bactérias e arqueias (Galvez et al., 2008; Wieland Brown et al., 2009; Piper et al., 2009). Elas são capazes de inibir o crescimento de outras bactérias (Cotter et al., 2012). Os terpenoides por sua vez constituem uma das maiores e mais diversa família de produtos naturais e possuem ampla aplicação devido as suas propriedades químicas (Bian et al., 2017; Schulz et al., 2017). Tem a função de comunicadores celulares, além de diminuir a competição entre as bactérias (Schwab et al., 2012; Abbas et al. 2017).

Outro agrupamento identificado através da mineração genômica foi glicolipídeos heterocísticos. Esses compostos são característicos de espécies fixadores de nitrogênio e apresentam elevado nível de informação quimiotaxonômica, permitindo assim identificar por exemplo a ordem Nostocales em nível de família ou gênero (Bauersachs et al., 2014). Outro agrupamento gênico identificado foi o relacionado a via NRPS, que são enzimas multimodulares sintetizadas por uma combinação de lipídios e aminoácidos, produzidos principalmente por bactérias e fungos, responsáveis pela síntese de vários compostos (WANG et al., 2014).

O agrupamento de nostoficina identificado no genoma da cepa GBBB01 se estende por 37,8 kb, três genes biossintéticas centrais, e mostrou 36% de similaridade com cluster de genes

biossintéticos de nostoficina de *Nostoc* sp. 152 (BGC0001029) (Fig. 8). Também foi identificada no genoma de *Nostoc* sp. ATCC 53789, *Nostoc* sp. CENA543 e *Nostoc punctiforme* PCC 73102, nesses genomas, apresentou 27% de similaridade. Sintetizada por um híbrido complexo enzimático PKS / NRPS (Fewer, et al., 2011). Esse metabólito foi considerado não tóxico contra bactérias (*S. aureus*, *Bacillus subtilis*, *E. coli*) e fungos (*Aspergillus niger*, *C. albicans*) e mostrou fraca citotoxicidade contra as células de leucemia linfocítica de camundongo (Fujii et al., 1999; Fewer, et al., 2011; Nowruzi; Porzani, 2020).

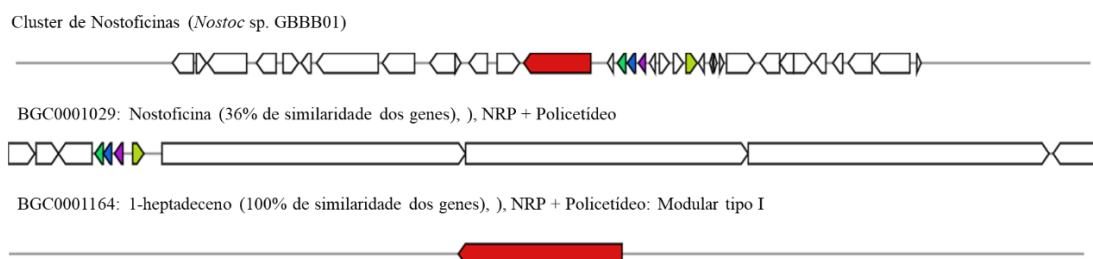


Fig.8. Agrupamento de Nostoficina comparado com outros agrupamentos de metabólitos similares realizado pelo KnowClusterBlast do antiSMASH.

Outro agrupamento gênico detectado no genoma da cepa foi de Microviridina k, com comprimento de 20,3 kb e dois genes centrais, apresentou 25% de similaridade, com cluster do gene biossintético de microviridina K de *Planktothrix agardhii* NIVA-CYA 126/8 (BGC0000594) (Fig. 9). Na análise dos demais genomas, este agrupamento foi apontado no genoma de *Nostoc* sp. NIES-2111, o qual mostrou 62% de similaridade com o cluster de referência. Esse composto é sintetizado ribossomicamente e modificados pós-tradução (Philmus et al., 2008; Zhong et al. 2020; Do Amaral et al., 2021). Estão entre os peptídeos mais promissores encontrados no filo das cianobactérias, identificados em uma variedade de gênero como *Microcystis*, *Nodularia*, *Planktothrix*, *Anabaena* e *Nostoc* (Ziemert et al., 2008; Do Amaral et al., 2021). Considerados potentes inibidores de protease, esse composto pode ser utilizado para diversos fins como para desenvolvimento de medicamentos e o controle de vetores (Do Amaral et al., 2021).

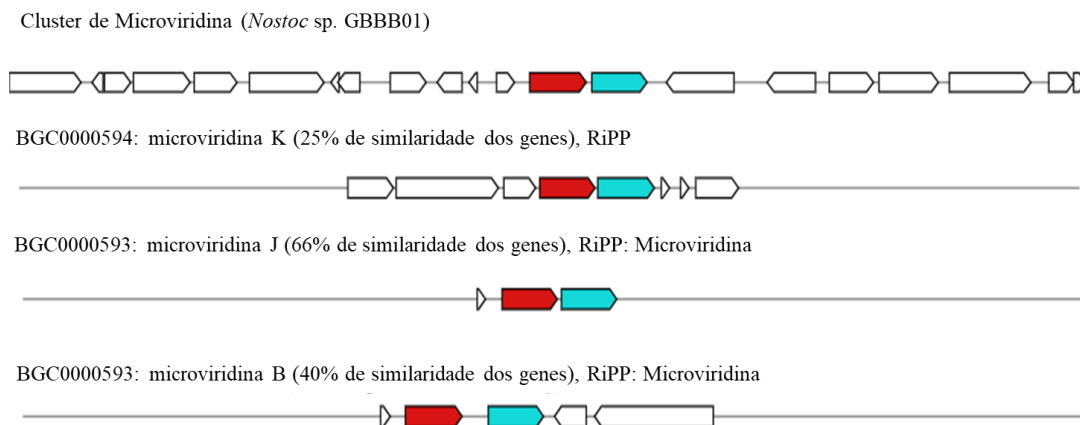


Fig.9. Agrupamento de Microviridina K comparado com outros agrupamentos de metabólicos similares realizado pelo KnowClusterBlast do. antiSMASH

O agrupamento gênico de Nostopeptolídeo A2 detectado no genoma da cepa GBBB01, se estende por 94 kb, com quatro genes biossintéticos centrais e apresentou 50% de similaridade com cluster de genes biossintéticos do nostopeptolídeo A2 de *Nostoc* sp. GSV224 (BGC0001028) (Fig. 10). Este agrupamento também foi apontado no genoma de *Nostoc* sp. NIES-2111 e *Nostoc punctiforme* PCC 73102, ambos com 100% de similaridade. O nostopeptolídeo em *Nostoc punctiforme* desempenha um papel importante na repressão do hormogônio e na diferenciação celular, além disso, diferentes concentrações de nostopeptolídeo liberado, apresentam efeitos completamente diferentes na diferenciação celular e na direção da motilidade (Liaimer et al.,2015).

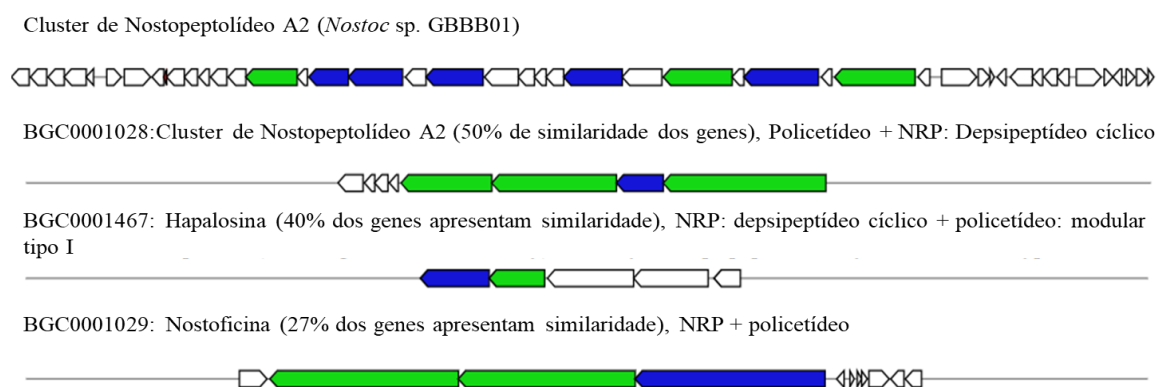


Fig.10. Agrupamento de Nostopeptolídeo A2 comparado com outros agrupamentos de metabólicos similares realizado pelo KnowClusterBlast do. antiSMASH.

Dentre os metabólitos secundários detectado somente no genoma da GBBB01, destaca-se agrupamento de Nosperina. Este cluster apresentou comprimento de 50 kb, compreendendo seis genes centrais e sete genes menores que codificam enzimas acessórios. A comparação do

cluster, mostrou 100% de similaridade dos genes com o cluster de nosperina de *Nostoc* sp. *Peltigera membranaceae* (BGC0001071) e cusperina de *Cuspidothrix issatschenkoi* CHARLIE-1 (BGC0001564) (Fig.11), indicando que esse agrupamento pode estar possivelmente envolvido na biossíntese desse metabólito. A nosperina é um membro distinto da família de compostos da Pederina, oriundo da simbiose do líquen *Peltigera membranacea* e da *Nostoc* sp. (Kampa et al., 2013). Os membros desta família possuem potente atividade antitumoral, causada pela inibição seletiva do ribossomo eucariótico (Kust, et al., 2018). Entretanto para o metabólito identificado na cepa GBBB01 existem poucos trabalhos conclusivos acerca da sua bioatividade, apenas sugere-se ter bioatividade semelhante ao demais membros da família (Kampa et al., 2013).

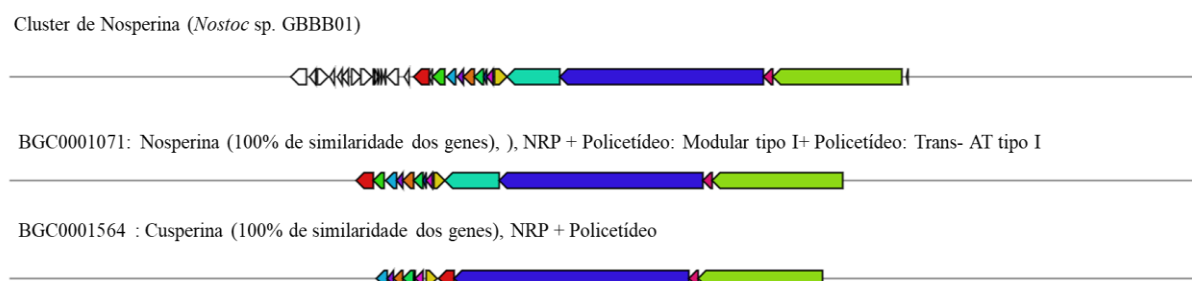


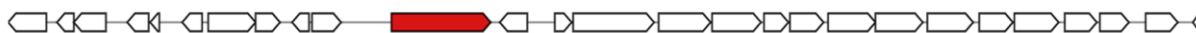
Fig.11. Agrupamento de Nosperina comparado com outros agrupamentos de metabólitos similares realizado pelo KnowClusterBlast do antiSMASH.

O cluster de fosfato também foi detectado somente no genoma da GBBB01, este é um composto importante em processos fundamentais do metabolismo dos seres vivos como armazenamento de energia e estruturação da membrana, e nas cianobactérias também desempenham papel fundamental no crescimento em ambientes com condições desfavoráveis de nutrientes (Esteves, 1998).

No genoma da GBBB01 também identificou um agrupamento gênico possivelmente envolvido na biossíntese da anabaenopeptina NZ857/ Nostamida A, com 32 kb, e um único gene central envolvido na biossíntese deste composto. Este cluster apresentou 100% de similaridade com cluster de genes biossintéticos de anabaenopeptina NZ857 de *Nostoc punctiforme* PCC 73102 (Fig.12). As anabaenopeptinas pertencem a uma família altamente diversa de hexapeptídeos cíclicos (Rouhiainen et al., 2010) amplamente distribuídos no filo das cianobactérias. Detectadas em vários gêneros como *Anabaena*, *Nostoc*, *Nodularia* e *Planktothrix* (Rouge et al., 2009; Christiansen et al., 2011; Rouhiainen et al., 2010; Koketsu et al., 2013). Já foram descritas 115 variantes de anabaenopeptina, que diferem na composição dos seus aminoácidos (Shishido et al., 2017). De acordo com a literatura esses compostos

apresentam várias atividades biológicas como inibição da proteína fosfatase, carboxipeptidases A e U e serina/protease (Murakami et al., 2000; Sano et al., 2001; Welker; Dohren,2006; Harms et al., 2016; Mazur-Marzec et al.,2018).

Cluster de Anabaenopepitina NZ857/Nostamida A (*Nostoc* sp. GBBB01)



BGC0001479:Cluster de Anabaenopepitina NZ857/Nostamida A (100% de similaridade dos genes), NRP

Fig.12. Agrupamento de Anabaenopepitina NZ857/ Nostamida A comparado com outros agrupamentos de metabólicos similares realizado pelo KnowClusterBlast do. antiSMASH.

A busca por regiões conservadas do domínio C realizada pelo servidor NaPDoS para o genoma da *Nostoc* sp. GBBB01 detectou 15 regiões conservadas de domínio C (Tabela suplementar S7). Quando comparadas com as regiões dos outros genomas, os domínios mais encontrados foram os relacionados às vias sirigomicinas, microcistinas, gramicidina, bleomicina, liquenicina e PKS/NRPS2, porém somente no genoma da cepa GBBB01 foram identificadas as vias gramicidina e liquenicina (Figura Suplementar S1A). Em relação a regiões conservadas de domínio KS, identificou-se 21 domínios (Tabela Suplementar S8). Quando comparado, os mais observados foram os pertencentes as vias eptilonas, ácido graxos poliinsaturado e síntese de ácidos graxos (Figura Suplementar S1B). No entanto as vias leinamicina, cerromicina, jamaicamida, caliqueamicina e tetronomicina, só foram observadas no genoma da GBBB01.

Dentre as vias possivelmente relacionada a biossíntese de compostos apontados pelo NaPDoS, destaca-se o domínio relacionado a biossíntese de microcistina, está é considerada a classe de cianotoxinas mais difundida e investigada (Bouaïcha et al., 2019; Meriluoto, 2017). Por colocarem em risco a saúde humana e animal, devido aos seus efeitos tóxicos (Pham; Utsumi, 2018). Outro domínio detectado foi relacionado a via de gramicidina, que é um biocomposto que apresenta atividade antibacteriana e antifúngica. Além disso, apresentou atividade antitumoral nas células do câncer gástrico, oferecendo a possibilidade de ser aplicada ao tratamento deste tipo de câncer (Chen et al., 2019). Também apontou domínio de regiões conservadas relacionado a biossíntese do composto bleomicina, conforme a literatura apresenta atividade antibacteriana e antitumoral de amplo espectro (Du et al., 2000; Chen et al.,2018). Outro domínio identificado foi relacionado a via de eptilonas, este por sua vez também são classificadas como um potencial atividade anticâncer que inibe o crescimento das células

cancerígenas humanas por meio da estabilização dos microtúbulos celulares (Tang et al., 2000; Molnár, et al., 2000; Altmann; Gaugaz; Schiess, 2011; Cheng; Huang, 2018). Além das vias relacionadas a biossíntese de jamaicamidas apresentam atividade de bloqueio dos canais de sódio e toxicidade em linhagem celular de câncer de pulmão humano, neuroblastoma em camundongos e peixes (Edwards et al., 2004).

Entretanto para considerar que os domínios possam estar associados à produção do composto ou de um composto semelhante, os valores de identidade precisam estar entre 85% e 90% (Ziemert et al., 2012). Fato que não foi observado em nenhuma das regiões conservadas C e KS detectados nos genomas das cianobactérias comparadas, onde estas apresentaram valores de identidade abaixo de 65%.

Para concluir, este é o primeiro registro de um genoma do gênero *Nostoc* do estado do Maranhão (Cerrado brasileiro) e demonstrou a presença de genes relacionados biossíntese de diversos compostos, que precisam ser melhores estudados acerca de se conhecer sua importância do ponto de saúde e ambiental. Vale ressaltar que o genoma de *Nostoc* GBBB01 engloba vários genes únicos, quando comparados com outras linhagens filogeneticamente próximas, sendo que muitos CDSs ainda não possuem funções bem caracterizadas e, portanto, revelam o alto potencial da cepa em produzir compostos bioativos relevantes para aplicações biotecnológicas.

REFERÊNCIAS

- ADAMS, David G. Heterocyst formation in cyanobacteria. **Current opinion in microbiology**, v. 3, n. 6, p. 618-624, 2000.
- ALBERTSEN, Mads et al. Genome sequences of rare, uncultured bacteria obtained by differential coverage binning of multiple metagenomes. **Nature biotechnology**, v. 31, n. 6, p. 533-538, 2013.
- ALNEBERG, Johannes et al. Binning metagenomic contigs by coverage and composition. **Nature methods**, v. 11, n. 11, p. 1144-1146, 2014.
- ALNEBERG, Johannes et al. CONCOCT: clustering contigs on coverage and composition. **arXiv preprint arXiv:1312.4038**, 2013.
- ALVARENGA, Danillo O.; FIORE, Marli F.; VARANI, Alessandro M. A metagenomic approach to cyanobacterial genomics. **Frontiers in microbiology**, v. 8, p. 809, 2017.
- ARNDT, D., Grant, J., Marcu, A., Sajed, T., Pon, A., Liang, Y., Wishart, D.S. PHASTER: a better, faster version of the PHAST phage search tool. **Nucleic Acids Res**, 2016.
- ARKIN, Adam P et al. KBase: The United States Department of Energy Systems Biology Knowledgebase. **Nature biotechnology**, v. 36, n.7, p.566-569, 2018.
- AZIZ, Ramy K et al. The RAST Server: rapid annotations using subsystems technology. **BMC genomics**, v. 9, n.75, 2008.
- BARRICK, Jeffrey E. et al. Genome evolution and adaptation in a long-term experiment with *Escherichia coli*. **Nature**, v. 461, n. 7268, p. 1243-1247, 2009.
- BATSALOVA, T. et al. In vitro cytotoxicity and antioxidative potential of *Nostoc microscopium* (Nostocales, cyanobacteria). **Toxicol Forensic Med Open J**, v.1, n.1, p. 9–17. 2016.
- BAUERSACHS, Thorsten et al. Distribution of long chain heterocyst glycolipids in N₂-fixing cyanobacteria of the order Stigonematales. **Phytochemistry**, v. 98, p.145-150, 2014.
- BAYLY, Carmen L.; YADAV, Vikramaditya G. Towards precision engineering of canonical polyketide synthase domains: recent advances and future prospects. **Molecules**, v. 22, n. 2, p. 235, 2017.
- BECHER, Paul G. et al. The cyanobacterial alkaloid nostocarboline: an inhibitor of acetylcholinesterase and trypsin. **Journal of Applied Phycology**, v. 21, n. 1, p. 103-110, 2009.
- BERG, Katri A. et al. High diversity of cultivable heterotrophic bacteria in association with cyanobacterial water blooms. **The ISME journal**, v. 3, n. 3, p. 314-325, 2009.
- BERMAN-FRANK, I. et al. Nitrogen fixation and photosynthetic oxygen evolution in cyanobacteria. **Research in Microbiology**, v.54, n.3, p. 157-164, 2003.

BIAN, Guangkai et al. Strategies for terpenoid overproduction and new terpenoid discovery. **Current opinion in biotechnology**, v.48, p.234–241. 2017.

BICUDO, CE de M.; MENEZES, M. Gêneros de algas de águas continentais do Brasil (chave para identificação e descrições). São Carlos. **RiMa.**, 502p, 2006.

BLIN, Kai et al. antiSMASH 5.0: updates to the secondary metabolite genome mining pipeline. **Nucleic acids research**, v. 47, n. W1, p. W81-W87, 2019.

BLIN, Kai et al. antiSMASH 5.0: updates to the secondary metabolite genome mining pipeline. **Nucleic acids research**, v. 47, n.W1, 2019.

BOETTGER, Daniela; HERTWECK, Christian. Molecular diversity sculpted by fungal PKS–NRPS hybrids. **ChemBioChem**, v. 14, n. 1, p. 28-42, 2013.

BOOPATHI, Thangavelu; KI, Jang-Seu. Impact of environmental factors on the regulation of cyanotoxin production. **Toxins**, v. 6, n. 7, p. 1951-1978, 2014.

BORNET E', FLAHAULT C. Revision des Nostocace´es he´te´rocyste´es contenues dans les principaux herbiers de France (quatrie`me et dernier fragment). *Ann Sci Nat Bot* ,v.7, p.177–262, 1888.

BRODDRICK, Jared T. et al. Unique attributes of cyanobacterial metabolism revealed by improved genome-scale metabolic modeling and essential gene analysis. **Proceedings of the National Academy of Sciences**, v. 113, n. 51, p. E8344-E8353, 2016.

CARMICHAEL, Wayne W. et al. Human fatalities from cyanobacteria: chemical and biological evidence for cyanotoxins. **Environmental health perspectives**, v. 109, n. 7, p. 663-668, 2001.

CARPINE, Roberta; SIEBER, Simon. Antibacterial and Antiviral Metabolites from Cyanobacteria: Their Application and their Impact on Human Health. **Current Research in Biotechnology**, v.3, p.65-81, 2021.

CARPINE, Roberta; SIEBER, Simon. Antibacterial and Antiviral Metabolites from Cyanobacteria: Their Application and their Impact on Human Health. **Current Research in Biotechnology**, 2021.

CHRISTENSEN, Victoria G.; KHAN, Eakalak. Freshwater neurotoxins and concerns for human, animal, and ecosystem health: A review of anatoxin-a and saxitoxin. **Science of the Total Environment**, v. 736, p. 139515, 2020.

CORBEL, S. et al. Cyanobacterial toxins: modes of actions, fate in aquatic and soil ecosystems, phytotoxicity and bioaccumulation in agricultural crops. **Chemosphere**, v.96, p.1–15, 2014.
Cotter, Paul D et al. “Bacteriocins - a viable alternative to antibiotics?.” *Nature reviews. Microbiology*, v. 11, n.2, p. 95-105, 2013.

COTTER, P. D., ROSS, R. P., HILL, C. Bacteriocins - a viable alternative to antibiotics?. *Nature reviews. Microbiology*, v.11, n.2, p.95–105, 2013.

COUVIN D et al. CRISPRCasFinder, an update of CRISRFinder, includes a 14 portable version, enhanced performance and integrates search for Cas proteins. *Nucleic Acids Res*, v.46, p.W246–W251, 2018.

DEMAY, Justine et al. Natural products from cyanobacteria: focus on beneficial activities. *Marine drugs*, v. 17, n. 6, p. 320, 2019.

DITTMANN, E et al. “Molecular biology of peptide and polyketide biosynthesis in cyanobacteria.” *Applied microbiology and biotechnology*, v.57,n.4,p.467-73, 2001.

DITTMANN, Elke et al. Natural Product Biosynthetic Diversity and Comparative Genomics of the Cyanobacteria. *Trends in microbiology*, vol. 23, n.10, p. 642-652, 2015.

DITTMANN, Elke; BÖRNER, Thomas. Contribuições genéticas para a avaliação de risco da microcistina no meio ambiente. *Toxicology and Applied Pharmacology*, v. 203, n. 3, p. 192-200, 2005.

DRÖGE, Johannes; MCHARDY, Alice C. Taxonomic binning of metagenome samples generated by next-generation sequencing technologies. *Briefings in bioinformatics*, v. 13, n. 6, p. 646-655, 2012.

DU, X. et al. The Diversity of Cyanobacterial Toxins on Structural Characterization, Distribution and Identification: A Systematic Review. *Toxins*, v.11, p:530, 2019.

FELSENSTEIN, J. Confidence limits on phylogenies: An approach using the bootstrap. *Evolution* v.39,p.783-791, 1985.

FEWER, David P. et al. Nostophycin biosynthesis is directed by a hybrid polyketide synthase-nonribosomal peptide synthetase in the toxic cyanobacterium *Nostoc* sp. strain 152. *Applied and environmental microbiology*, v. 77, n. 22, p. 8034-8040, 2011.

FIDOR, Anna; KONKEL, Robert; MAZUR-MARZEC, Hanna. Bioactive peptides produced by cyanobacteria of the genus *Nostoc*: A review. *Marine drugs*, v. 17, n. 10, p. 561, 2019.

FUJII, Kiyonaga et al. Nostophycin, a novel cyclic peptide from the toxic cyanobacterium *Nostoc* sp. 152. *The Journal of Organic Chemistry*, v. 64, n. 16, p. 5777-5782, 1999.

GADEMANN, Karl; PORTMANN, Cyril. Secondary metabolites from cyanobacteria: complex structures and powerful bioactivities. *Current Organic Chemistry*, v. 12, n. 4, p. 326-341, 2008.

GAGUNASHVILI, Andrey N.; ANDRESSON, Olafur S. Distinctive characters of *Nostoc* genomes in cyanolichens. *BMC genomics*, v. 19, n. 1, p. 1-18, 2018.

GRÜNEWALD, Jan; MARAHIEL, Mohamed A. Chemoenzymatic and template- 49
synthesis of bioactive macrocyclic peptides. *Microbiology and molecular biology rev*
70, n. 1, p. 121-146, 2006.

HALL, T., Biosciences, I.; Carlsbad, C. BioEdit: an important software for molecular biology. *GERF Bull Biosci*, v.2, n.1, p.60-61. 2011.

HARMS, Henrik et al. Discovery of anabaenopeptin 679 from freshwater algal bloom material: Insights into the structure-activity relationship of anabaenopeptin protease inhibitors. **Bioorganic & medicinal chemistry letters**, v. 26, n. 20, p.4960-4965. 2016.

HEMSCHEIDT, Thomas K. Microviridin biosynthesis. **Methods in enzymology**, v. 516, p. 25-35, 2012.

HROUZEK, Pavel et al. Description of the cyanobacterial genus *Desmonostoc* gen. nov. including *D. muscorum* comb. nov. as a distinct, phylogenetically coherent taxon related to the genus *Nostoc*. **Fottea**, v. 13, n. 2, p. 201-213, 2013.

HUANG, I.-Shuo; ZIMBA, Paul V. Cyanobacterial bioactive metabolites. A review of their chemistry and biology. **Harmful algae**, v. 86, p. 139-209, 2019.

HWANG, Soonkyu et al. Repurposing Modular Polyketide Synthases and Non-ribosomal Peptide Synthetases for Novel Chemical Biosynthesis. **Frontiers in molecular biosciences**, v. 7, n.87, p.15, 2020.

JEANNE-JULIEN, Louis et al. Synthesis of a Tiacumicin B Protected Aglycone. **Organic letters**, v. 19, n.15, p.4006-4009, 2017.

JEONG, Yujin et al. Current Status and Future Strategies to Increase Secondary Metabolite Production from Cyanobacteria. **Microorganisms**, v.8, n.12, p.1849, 2020.

JOKELA, Jouni et al. A novel cyanobacterial nostocyclopeptide is a potent antitoxin against microcystins. **ChemBioChem**, v. 11, n. 11, p. 1594-1599, 2010.

JOKELA, Jouni et al. Production of high amounts of hepatotoxin nodularin and new protease inhibitors pseudospumigins by the Brazilian benthic *Nostoc* sp. CENA543. **Frontiers in microbiology**, v. 8, p. 1963, 2017.

KANG, Dongwan D et al. MetaBAT 2: an adaptive binning algorithm for robust and efficient genome reconstruction from metagenome assemblies. **PeerJ**, v. 7, p.7359, 2019.

KAPLAN-LEVY, Ruth N. et al. Akinetes: dormant cells of cyanobacteria. **Dormancy and resistance in harsh environments**, p. 5-27, 2010.

KEHR, Jan-Christoph; PICCHI, Douglas Gatte; DITTMANN, Elke. Natural product biosyntheses in cyanobacteria: a treasure trove of unique enzymes. **Beilstein journal of organic chemistry**, v. 7, n. 1, p. 1622-1635, 2011.

KIMURA M. A simple method for estimating evolutionary rate of base substitutions through comparative studies of nucleotide sequences. **Journal of Molecular Evolution**, v.16, n.2, p.111-120, 1980.

KITAGAKI, Shinji et al. Planar Chiral [2.2] Paracyclophane-Based Bisoxazoline Ligands: Design, Synthesis, and Use in Cu-Catalyzed Inter- and Intramolecular Asymmetric O–H Insertion Reactions. **Chemical and Pharmaceutical Bulletin**, v. 66, n. 10, p. 1006-1014, 2018.

KOKETSU, Kento et al. Identification of homophenylalanine biosynthetic genes from the cyanobacterium *Nostoc punctiforme* PCC73102 and application to its microbial production by *Escherichia coli*. **Applied and environmental microbiology**, v. 79, n.7, p. 2201-2208, 2013.

KOMÁREK, J. 19/3: Cyanoprokaryota; Teil 3: Heterocytous genera. **Berlin [etc.]: Springer**, 2013.

KOMÁREK, J. A polyphasic approach for the taxonomy of cyanobacteria: principles and applications. **European Journal of Phycology**, v. 51, n.3, p.346–353, 2016.

KOMÁREK, J. Review of the cyanobacterial genera implying planktic species after recent taxonomic revisions according to polyphasic methods: state as of 2014. **Hydrobiologia**, v. 764, n. 1, p. 259-270, 2015.

KOMÁREK, Jiří et al. Quo vadis, taxonomy of cyanobacteria (2019). **Fottea**, v. 20, n. 1, p. 104-110, 2020.

KOMÁREK, Jiří. Recent changes (2008) in cyanobacteria taxonomy based on a combination of molecular background with phenotype and ecological consequences (genus and species concept). **Hydrobiologia**, v. 639, n. 1, p. 245-259, 2010.

KOONIN, Eugene V. Evolution of genome architecture. **The international journal of biochemistry & cell biology**, v. 41, n. 2, p. 298-306, 2009.

KUMAR, Jay et al. Cyanobacteria: applications in biotechnology. In: **Cyanobacteria**. Academic Press, p. 327-346, 2019.

KUMAR, Sudhir et al. MEGA X: Molecular Evolutionary Genetics Analysis across Computing Platforms. **Molecular biology and evolution**, v. 35, n.6, p.1547-1549, 2018.

KUNIN, Victor et al. A bioinformatician's guide to metagenomics. **Microbiology and molecular biology reviews**, v. 72, n. 4, p. 557-578, 2008.

KUST, Andreja et al. A new microcystin producing *Nostoc* strain discovered in broad toxicological screening of non-planktic Nostocaceae (cyanobacteria). **Toxicon: official journal of the International Society on Toxinology**, v.150, p. 66-73, 2018.

KVÍDEROVÁ, J. et al. **Ecophysiology of Cyanobacteria in the Polar Regions**. *Cyanobacteria*, p.277–302, 2019.

LARSSON, John; NYLANDER, Johan AA; BERGMAN, Birgitta. Genome fluctuations in cyanobacteria reflect evolutionary, developmental and adaptive traits. **BMC evolutionary biology**, v. 11, n. 1, p. 1-21, 2011.

LATYSHEVA, Natasha et al. **The evolution of nitrogen fixation in cyanobacteria**. **Bioinformatics**, v. 28, n. 5, p. 603-606, 2012.

LAUGHINGHOUSE, Haywood Dail et al. Biology and management of *Nostoc* (cyanobacteria) in nurseries and greenhouses. **EDIS**, v. 2019, n. 2, 2019.

LIMA, Alex Ranieri Jerônimo et al. Draft genome sequence of the Brazilian Cyanobium sp. strain CACIAM 14. **Genome announcements**, v. 2, n. 4, 2014.

LIMA, R.N.; CÂMARA, R.J.B.; PÃOZINHO, F.C. O geoturismo como segmento diferencial para o polo turístico da Chapada das Mesas: estudo de viabilidade de práticas geoturísticas no Complexo de Pedra Caída em Carolina (MA). Anais do IX Congresso Nacional de Ecoturismo e do V Encontro Interdisciplinar de Turismo em Unidades de Conservação. **Revista Brasileira de Ecoturismo**, São Paulo, v.6, n.4, pp.103-120. 2013.

LIU, Liwei et al. Nostosins, trypsin inhibitors isolated from the terrestrial cyanobacterium *Nostoc* sp. strain FSN. **Journal of natural products**, v. 77, n. 8, p. 1784-1790, 2014.

MANDE, Sharmila S.; MOHAMMED, Monzoorul Haque; GHOSH, Tarini Shankar. Classification of metagenomic sequences: methods and challenges. **Briefings in bioinformatics**, v. 13, n. 6, p. 669-681, 2012.

MARQUES, A. N. et al. Abordagens sobre o território e a territorialidade das comunidades sertanejas do cerrado da Chapada Das Mesas / MA. **Revista de geografia agrária**.v.9, n.19. p 472-507, 2014.

MASCHEK, J. Alan; BAKER, Bill J. The chemistry of algal secondary metabolism. In: **Algal chemical ecology**. Springer, Berlin, Heidelberg, p. 1-24, 2008.

MASCHEK, J. Alan; BAKER, Bill J. The chemistry of algal secondary metabolism. In: **Algal chemical ecology**. Springer, Berlin, Heidelberg, 2008. p. 1-24.

MAZARD, S. et al. Tiny Microbes with a Big Impact: The Role of Cyanobacteria and Their Metabolites in Shaping Our Future. *Marine drugs*, 14: 97. 2016.

MAZUR-MARZEC, Hanna et al. Cyanopeptolins with Trypsin and Chymotrypsin Inhibitory Activity from the Cyanobacterium *Nostoc edaphicum* CCNP1411. **Marine drugs**, v. 16, n. 7, 2018.

MEEKS, John C. et al. Cellular differentiation in the cyanobacterium *Nostoc punctiforme*. **Archives of microbiology**, v. 178, n. 6, p. 395-403, 2002.

MEIER-KOLTHOFF et al. Genome sequence-based species delimitation with confidence intervals and improved distance functions. **BMC Bioinformatics**, v.14, n.60, 2013.

MEIER-KOLTHOFF, J. P. et al. Genome sequence-based species delimitation with confidence intervals and improved distance functions. **BMC Bioinformatics**,v.4, p.60, 2013.

MOREIRA, L. M. Ciências genômicas: fundamentos e aplicações. In: Prosdócimi, F. & Moreira, L. M. **Genômica comparativa**. Ribeirão Preto: Sociedade Brasileira de Genética, v. 1, p. 81- 100, 2015.

MOYA, Andrés et al. Driven progressive evolution of genome sequence complexity in Cyanobacteria. **Scientific reports**, v. 10, n. 1, p. 1-14, 2020.

MURAKAMI, M et al. New anabaenopeptins, potent carboxypeptidase-A inhibitors from the cyanobacterium *Aphanizomenon flos-aquae*. **Journal of natural products** , v.63, n. 9, p.1280-1282, 2000.

NIU, Siwen et al. “Characterization of a sugar-O-methyltransferase TiaS5 affords new Tiacumicin analogues with improved antibacterial properties and reveals substrate promiscuity.” *Chembiochem : a European journal of chemical biology*, v. 12, n.11, p.1740-8, 2011.

NOWRUZI, B. et al. Nostoc cyanobacteria species: a new and rich source of novel bioactive compounds with pharmaceutical potential. **Journal of Pharmaceutical Health Services Research**, v.9, p. 5-12, 2018.

NOWRUZI, Bahareh; PORZANI, Samaneh Jafari. Toxic compounds produced by cyanobacteria belonging to several species of the order Nostocales: A review. **Journal of Applied Toxicology**, v. 41, n. 4, p. 510-548, 2021.

NURK, Sergey et al. metaSPAdes: a new versatile metagenomic assembler. **Genome research**, v. 27, n.5, p.824-834. 2017.

O'CONNOR, Sarah E. Engineering of secondary metabolism. **Annual review of genetics**, v. 49, p. 71-94, 2015.

OVERBEEK, Ross et al. The SEED and the Rapid Annotation of microbial genomes using Subsystems Technology (RAST). **Nucleic acids research**, v. 42, 2014.

PIPER, Clare et al. Discovery of medically significant lantibiotics. **Current drug discovery Technologies**, v. 6, n. 1, p.1-18. 2009.

PRABHA, Ratna et al. Functional profiling of cyanobacterial genomes and its role in ecological adaptations. **Genomics Data**, v. 9, p. 89-94, 2016.

RAJANIEMI, Pirjo et al. Phylogenetic and morphological evaluation of the genera *Anabaena*, *Aphanizomenon*, *Trichormus* and *Nostoc* (Nostocales, Cyanobacteria). **International Journal of Systematic and Evolutionary Microbiology**, v. 55, n. 1, p. 11-26, 2005.

RIMSHA, Rizvi et al. Bioactive substances of cyanobacteria (*Nostoc muscorum*): a review. **International Journal of Pharma Sciences and Research**, v. 5, n. 07, p. 320-322, 2014.

RODGERS, K. J.; MAIN, B. J.; SAMARDZIC, K. Cyanobacterial Neurotoxins: Their Occurrence and Mechanisms of Toxicity. *Neurotoxicityresearch*, v.33,n.1, p:168–177. 2018. Disponível me < <https://doi.org/10.1007/s12640-017-9757-2>>

ROUHIAINEN, Leo et al. Two alternative starter modules for the non-ribosomal biosynthesis of specific anabaenopeptin variants in *Anabaena* (Cyanobacteria). **Chemistry & biology**, v. 17, n.3, p. 265-273, 2010.

ROUNGE, Trine B et al. A genome-wide analysis of nonribosomal peptide synthetase gene clusters and their peptides in a *Planktothrix rubescens* strain. **BMC genomics**, vol. 10, n.396, 2009.

ROY-LACHAPELLE, Audrey et al. Analysis of multiclass cyanotoxins (microcystins, anabaenopeptins, cylindrospermopsin and anatoxins) in lake waters using on-line SPE liquid chromatography high-resolution Orbitrap mass spectrometry. **Analytical Methods**, v. 11, n. 41, p. 5289-5300, 2019.

SAITOU N. e Nei M. The neighbor-joining method: A new method for reconstructing phylogenetic trees. **Molecular Biology and Evolution**, v.4, n.4, p.406-425. 1987.

SANO, T et al. Isolation of new protein phosphatase inhibitors from two cyanobacteria species, *Planktothrix* spp. **Journal of natural products**, v. 64, n.8, p.1052-5. 2001.

SANT'ANNA C.L.; AZEVEDO, M.T.P.; AGUJARO, L.F.; CARVALHO, M.C.; CARVALHO, L.R.; SOUZA, R.C.R. Manual ilustrado para identificação e contagem de cianobactérias planctônicas de águas continentais brasileiras. Rio de Janeiro: Interciência; p53. 2006.

SARAF, Aniket et al. Description of two new species of *Aliinostoc* and one new species of *Desmonostoc* from India based on the Polyphasic Approach and reclassification of *Nostoc punensis* to *Desmonostoc punense* comb. nov. **FEMS microbiology letters**, v. 365, n. 24, p. fny272, 2018.

SCHULZ-BOHM, K., MARTÍN-SÁNCHEZ, L.; GARBEVA, P. Microbial Volatiles: Small Molecules with an Important Role in Intra- and Inter-Kingdom Interactions. **Frontiers in microbiology**, v.8, p.2484. 2017.

SCHULZ-BOHM, Kristin et al. Microbial Volatiles: Small Molecules with an Important Role in Intra- and Inter-Kingdom Interactions. **Frontiers in microbiology**, v. 8, 2017.

SCHWARZER, Dirk; FINKING, Robert; MARAHIEL, Mohamed A. Nonribosomal peptides: from genes to products. **Natural product reports**, v. 20, n. 3, p. 275-287, 2003.

SHAH, Sayed Asmat Ali et al. Structural diversity, biological properties and applications of natural products from cyanobacteria. A review. **Marine drugs**, v. 15, n. 11, p. 354, 2017.

SHIH, Patrick M. et al. Improving the coverage of the cyanobacterial phylum using diversity-driven genome sequencing. **Proceedings of the National Academy of Sciences**, v. 110, n. 3, p. 1053-1058, 2013.

SHISHIDO, Tânia K et al. "Simultaneous Production of Anabaenopeptins and Namalides by the Cyanobacterium *Nostoc* sp. CENA543. **ACS chemical biology**, v. 12, n. 11, p. 2746-2755, 2017.

SHISHIDO, Tania Keiko et al. Dereplication of Natural Products with Antimicrobial and Anticancer Activity from Brazilian Cyanobacteria. **Toxins** v.12, n.1, p.12.-24, 2019.

SILVA, M. L. et al. Percepção ambiental dos moradores do Parque Nacional da Chapada das Mesas, no domínio fitogeográfico do Cerrado Brasileiro. **Revista Espacios**, v. 38, n. 22, p. 33, 2016.

SILVA, M.L. A.; ARAÚJO, M. F. V.; CONCEICAO, G. M.. Identidade e modo de vida dos moradores do Parque Nacional da Chapada das Mesas, Maranhão, Brasil. **Biota Amazônia (Biote Amazonie, Biota Amazonia, Amazonian Biota)**, v. 7, n. 4, p. 41-47, 2017.

SIMM, Stefan et al. The composition of the global and feature specific cyanobacterial core-genomes. **Frontiers in microbiology**, v. 6, p. 219, 2015.

SINGH, Jay Shankar; PANDEY, Vimal Chandra; SINGH, Davinder P. Efficient soil microorganisms: a new dimension for sustainable agriculture and environmental development. **Agriculture, ecosystems & environment**, v. 140, n. 3-4, p. 339-353, 2011.

SINGH, Prashant et al. New species of Nostoc (cyanobacteria) isolated from Pune, India, using morphological, ecological and molecular attributes. **Plant Systematics and Evolution**, v. 302, n. 10, p. 1381-1394, 2016.

SINGH, Rachana et al. Uncovering potential applications of cyanobacteria and algal metabolites in biology, agriculture and medicine: current status and future prospects. **Frontiers in microbiology**, v. 8, p. 515, 2017.

SIVONEN, Kaarina; JONES, Gary. Cyanobacterial toxins. **Toxic cyanobacteria in water: a guide to their public health consequences, monitoring and management**, v. 1, p. 43-112, 1999.

SOUSA, R. C. **Caracterização polifásica de microrganismos fotossintetizantes da Chapada das Mesas**. Dissertação (mestrado em saúde e ambiente). Programa de pós-graduação em saúde e ambiente, Universidade Federal do Maranhão, 2019. Disponível em <<https://tedebc.ufma.br/jspui/handle/tede/3120>>

SOUZA, F. T.; KOERNER, T. C.; CHLAD, R. A data-based model for predicting wildfires in Chapada das Mesas National Park in the State of Maranhão. **Environmental Earth Sciences**, v. 74, n. 4, p.3603-3611, 2015.

SURADKAR, Archana et al. Nostoc thermotolerans sp. nov., a soil-dwelling species of Nostoc (Cyanobacteria). **International journal of systematic and evolutionary microbiology**, v. 67, n. 5, p. 1296-1305, 2017.

TATUSOVA, Tatiana et al. NCBI prokaryotic genome annotation pipeline. **Nucleic acids research**, v. 44, n.14, p.6614-24. 2016.

TEKAIA, F. Inferring orthologs: Open question and perspectives. **Genomics Insights**, v.9, p.17-28, 2016.

TENAILLON, Olivier et al. Tempo and mode of genome evolution in a 50,000-generation experiment. **Nature**, v. 536, n. 7615, p. 165-170, 2016.

THUAN, N. H. et al. Recent advances in exploration and biotechnological production of

bioactive compounds in three cyanobacterial genera: *Nostoc*, *Lyngbya*, and *Microcystis*. **Frontiers in chemistry**, v.7, p.604, 2019.

URITSKIY, Gherman V et al. MetaWRAP—a flexible pipeline for genome-resolved metagenomic data analysis. **Microbiome**, v. 6, n.1, p.158, 2018.

VIJAYAKUMAR, Subramaniyan; MENAKHA, Muniraj. Pharmaceutical applications of cyanobacteria—A review. **Journal of Acute Medicine**, v. 5, n. 1, p. 15-23, 2015.

WANG, H. et al. Genome mining demonstrates the widespread occurrence of gene clusters encoding bacteriocins in cyanobacteria. **PLoS one**, v.6, p. 22384, 2011.

WANG, Hao et al. Genome-derived insights into the biology of the hepatotoxic bloom-forming cyanobacterium *Anabaena* sp. strain 90. **BMC genomics**, v. 13, n. 1, p. 1-17, 2012.

WATANABE, M. M.; Tanabe, Y. Biology and Industrial Potential of *Botryococcus braunii*. In: RICHMOND, A.; HU, Q. (Ed.). **Handbook of Microalgal Culture**. Oxford, Reino Unido: John Wiley & Sons Ltda, p. 369–387. 2013.

WATERBURY, John B. The cyanobacteria— isolation, purification and identification. **The prokaryotes**, v. 4, p. 1053-1073, 2006.

WEISS, Christine et al. Cryptophycins: Cytotoxic cyclodepsipeptides with potential for tumor targeting. **Journal of Peptide Science**, v. 23, n. 7-8, p. 514-531, 2017.

WELKER M, VON DÖHREN H. Cyanobacterial peptides—nature's own combinatorial biosynthesis. **FEMS Microbiol Rev.** v. 30, n.4, p:530-563, 2006.

WELKER, Martin; VON DÖHREN, Hans. Cyanobacterial peptides—nature's own combinatorial biosynthesis. **FEMS microbiology reviews**, v. 30, n. 4, p. 530-563, 2006.

WEN, Z.Y.; CHEN, F. Perfusion culture of the diatom *Nitzschia laevis* for ultra-high yield of eicosapentaenoic acid. **Process Biochemistry**, v.38, n.4, p.523–529, 2002.

WHITTON, B. A, POTTS, M. Introduction to the cyanobacteria. In: Whitton BA, Potts M (eds) *Ecology of cyanobacteria: their diversity in time and space*. Kluwer Academic Publishers, Dordrecht, p 1–11, 689, 2000.

WHITTON, Brian A.; POTTS, Malcolm. Introduction to the cyanobacteria. In: **Ecology of Cyanobacteria II**. Springer, Dordrecht, p. 1-13, 2012.

WIESE, Maria et al. Neurotoxic alkaloids: saxitoxin and its analogs. **Marine drugs**, v. 8, n. 7, p. 2185-2211, 2010.

WU, Yu-Wei et al. MaxBin 2.0: an automated binning algorithm to recover genomes from multiple metagenomic datasets. **Bioinformatics**, v. 32, p. 605-607. 2016.

XIAO, Yi et al. Characterization of tiacumicin B biosynthetic gene cluster affording diversified tiacumicin analogues and revealing a tailoring dihalogenase. **Journal of the American Chemical Society**, v. 133, n.4, p. 1092-105, 2011.

XU, Ling et al. OrthoVenn2: a web server for whole-genome comparison and annotation of orthologous clusters across multiple species. **Nucleic acids research**, v. 47, n.W1, p.W52-W58. 2019.

YU, Dayu et al. Type III polyketide synthases in natural product biosynthesis. **IUBMB life**, v. 64, n. 4, p. 285-295, 2012.

YU, Haifeng; JIA, Shiru; DAI, Yujie. Growth characteristics of the cyanobacterium *Nostoc flagelliforme* in photoautotrophic, mixotrophic and heterotrophic cultivation. **Journal of Applied Phycology**, v. 21, n. 1, p. 127-133, 2009.

ZAHRA Zahra, et al. Cyanobacteria: Review of Current Potentials and Applications. **Environments** . v.7, n.2, p.13, 2020.

ZEHR, Jonathan P. et al. Globally distributed uncultivated oceanic N₂-fixing cyanobacteria lack oxygenic photosystem II. **science**, v. 322, n. 5904, p. 1110-1112, 2008.

ZERVOU, Sevasti-Kiriaki et al. New microginins from cyanobacteria of Greek freshwaters. **Chemosphere**, v. 248, p. 125961, 2020.

ZHANG, Yi et al. A distributive peptide cyclase processes multiple microviridin core peptides within a single polypeptide substrate. **Nature communications**, v. 9, n. 1, p. 1-10, 2018.

ZHOU, Y. et al. PHAST: a fast phage search tool. *Nucleic Acids Res.*, 39, p.W347–W352. 2011.

ZIEMERT, Nadine et al. The natural product domain seeker NaPDoS: a phylogeny based bioinformatic tool to classify secondary metabolite gene diversity. **PloS one**, v. 7, n. 3, p. e34064, 2012.

ZIEMERT, Nadine et al. The natural product domain seeker NaPDoS: a phylogeny based bioinformatic tool to classify secondary metabolite gene diversity. **PloS one**, v. 7, n.3, 2012.

ZURAWELL, Ronald W. et al. Hepatotoxic cyanobacteria: a review of the biological importance of microcystins in freshwater environments. **Journal of Toxicology and Environmental Health, Part B**, v. 8, n. 1, p. 1-37, 2005.

ZURAWELL, Ronald W.; HOLMES, Charles FB; PREPAS, Ellie E. Elimination of the cyanobacterial hepatotoxin microcystin from the freshwater pulmonate snail *Lymnaea stagnalis jugularis* (Say). **Journal of Toxicology and Environmental Health, Part A**, v. 69, n. 4, p. 303-318, 2006.

MATERIAL SUPLEMENTA

Tabela S1- Os genomas recuperados da cultura não axênica pelo sequenciamento *do Illumina MiSeq*.

GBBB01	Compleitude	Contaminação	N50	Tamanho	Taxonomia (GTDB-TK)	Classificação (NCBI)
bin_3	71.71	1.739	6479 pb	1,6 Mb	Alphaproteobacteria (c)	
bin_5	99.55	2.111	154223 pb	8,1 Mb	<i>Nostoc</i> (g)*	Nostoc (g)*
bin_6	100.00	0.0	107287 pb	3,8 Mb	<i>Leptospira</i> (g)	
bin_7	72.24	0.0	450720 pb	2 Mb	<i>Brevundimonas</i> (g)	
bin_9	98.90	3.296	211785 pb	4,5Mb	Gemmatimonadaceae (f)	
bin_10	99.29	0.9	149446 pb	4 Mb	Hydrogenophaga (g)	

Tabela S2- Sequências identificadas de CRISPR/CAS no genoma da *Nostoc* sp. GBBB01.

Elemento	ID CRISPR / Tipo de Cas	Início	Fim	Espaçador / gene	Repetição consenso / Genes Cas	Direção	Nível de Evidência
CRISPR	2604138_3_peg_95_1	551	648	1 1	ATGAATGGCTACCAGAGATCCAGTTGA	ND	11
CRISPR	2604138_3_peg_472_1	193	208	11	AAAAAAGTTGGGGTAAACAGGCGGAT CTTGGTGGATCAGGCGGATCA	ND	11
CRISPR	2604138_3_peg_756_1	475	680	4	GGTACTGGTAGCGGCATTGGTAGCGG	ND	2
CRISPR	2604138_3_peg_917_1	3	116	11	GTTTCCATTAATTCCACTTCTCAAAG AGAAGTGGC	ND	11
CRISPR	2604138_3_peg_113_8_1	16	126	11	GTTTCCATTAATTCCACTTCTCAGAAG AGAAGCGGC	ND	11
CRISPR	2604138_3_peg_133_5_1	476	587	11	CGCGCGGTAATCCCGAAGAGACGATG CGTACAGCGGC	ND	11
CRISPR	2604138_3_peg_149_8_1	244	350	11	GAAGAGTTCCAAGATTTCTCTACATTC TCGGAAGAATCAGA	ND	11
CRISPR	2604138_3_peg_225_6_1	51	219	2	AGGTTTCCACAGCCCTACCGTTAG	ND	11
CRISPR	2604138_3_peg_234_5_1	391	489	11	GGCGGCATCTACAACATATGATGG	ND	11
CRISPR	2604138_3_peg_234_5_2	1474	1652	2	GGCGGCATCTACAACATATGATGG	ND	11
CRISPR	2604138_3_peg_281_1_1	2279	2424	11	CCGACACCAAACCTTCTCCTCAAACAAC GCGCTACCCAGGATGAT	ND	11
CRISPR	2604138_3_peg_295_3_1	321	413	11	TAATAGCTCCGGTTCTAGATCTTC	ND	11
CRISPR	2604138_3_peg_345_5_1	868	962	11	ATAGTTAATAGTGGCGACGCAAT	ND	11
CRISPR	2604138_3_peg_420_6_1	505	621	11	CAAGTGGTAGACGCACTGATGTTGTAT CGTAG	ND	11
CRISPR	2604138_3_peg_445_2_1	3309	3405	1 1	AACTTATATAAACCCAGGGTAGTGGTGT GCAG	ND	11

CRISPR	2604138_3_peg_542 0_1	291	384	11	ACCTCCAGTACCACCACAACCAGGTGC T	ND	11
CRISPR	2604138_3_peg_575 6_1	46.	156	11	AGTTTCAATCCCGTTGCCGGGTTTCATT TAATTGAAAGT	ND	11
CRISPR	2604138_3_peg_575 7_1	44	153	11	GTTTCAATCCCGTTGCCGGGTTTCATTT AATTGA	ND	11
CRISPR	2604138_3_peg_577 1_1	38.	138	11	TGTTTCAATCCCGTTACCGGGTTTCATT TAA	ND	11
CRISPR	2604138_3_peg_590 3_1	2488	2630	2	AATCTTCAACAGGCTAATTTAAC	ND	11
CRISPR	2604138_3_peg_620 6_1	3	260	3	GTTTCCATTAATTCAACTTCCGAAGAA GCTAAAAG	ND	11
CRISPR	2604138_3_peg_628 5_1	439	518	11	TCTGCACAACAAGCAACCATCGCTCG	ND	11
CRISPR	2604138_3_peg_682 7_1	2620	2767	2	GGTTCTGGCAACGATAGCTTGAGTG	ND	11
CRISPR	2604138_3_peg_683 4_1	978	1062	11	TGTTGACTATTCAGATGGCAATAAC	ND	11
CRISPR	2604138_3_peg_704 1_1	406	514	11	GGTAATCGTCCTCAAGAAATTGCCCAA GCT	ND	11
CRISPR	2604138_3_peg_704 1_2	697	804	11	GGTAATCGTCCTCAAGAAATTGCCCAA GCT	ND	11
CRISPR	2604138_3_peg_718 7_1	45	153	11	CTTTACCGATTCCCTGAAATCGGAATT AATGGAAAC	ND	11
CRISPR	2604138_3_peg_725 8_1	2565	2675	11	CCTTGCTACACAAGCTGTTGAGGGTGT TGATCT	ND	11
CRISPR	2604138_3_peg_728 8_1	280	387	11	TTTGATATACATGGGCAAGACTAC	ND	11
CRISPR	2604138_3_peg_777 3_1	29	286	3	CGTTTCCATTAATTCGCCTCCAAATTGT TATTGGAGAC	ND	11

Tabela S3- Análise de identidade do gene 16S rRNA da *Nostoc* sp. GBBB01 gerados pelo BLASTn do NCBI.

Linagem	BLASTn - NCBI	Cobertura (%)	Identidade (%)	E value	Nº de acesso
	<i>Desmonostoc muscorum</i> TAU-MAC 0699	99%	99,67%	0,0	MN062666.1
	<i>Nostoc linckia</i> CALU 129	99%	99,67%	0,0	MG596748.1
<i>Nostoc</i> sp. GBBB01	<i>Desmonostoc</i> sp. PCC 8306	99%	99,67%	0,0	HG004584.1

Tabela S4: Distribuição dos principais genes ortólogos entre os domínios Processo Biológico, Função Molecular e Componente celular, gerados pelo OrthoVenn2.

Domínio Processo Biológico	Contagem de genes	Domínio Função Molecular	Contagem de genes	Domínio Componente celular	Contagem de genes
Processo metabólico (GO: 0008152)	710	Atividade de hidrolase (GO:0016787)	74	Membrana (GO:0016020)	58
Processo biológico (GO: 0008150)	719	Atividade de transferase (GO:0016740)	66	Parte da célula (GO:0044464)	56
Processo metabólico celular (GO: 0044237)	561	Ligação de íons (GO: 0043167)	45	Intracelular (GO: 0005622)	12
Processo metabólico do composto de nitrogênio (GO: 0006807)	415	Atividade oxidoredutase (GO: 0016491)	49	Componente celular (GO:0005575)	12
Processo celular (GO: 0009987)	402	Função_molecular (GO: 0003674)	45	Tilacóide (GO: 0009579)	10
Processo metabólico primário (GO: 0044238)	297	Atividade do transportador (GO: 0005215)	40	Plastídio (GO:0009536)	11

Processo metabólico heterociclo (GO: 0046483)	253	Atividade de peptidase (GO: 0008233)	33	Organela (GO:0043226)	3
Processo metabólico de compostos aromáticos celulares (GO: 0006725)	233	Ligação de ácido nucleico (GO: 0003676)	32	Organela intracelular (GO:0043229)	3
Processo metabólico da macromolécula (GO: 0043170)	229	Ligação de cofator (GO: 0048037)	18	Espaço periplásmico (GO:0042597)	3
Processo metabólico de ácido orgânico (GO: 0006082)	205	Ligação de nucleotídeo (GO: 0000166)	17		
Total	4024		419		168

Tabela S5: Distribuição dos principais genes únicos identificados pelo OrthoVenn2 nos genomas.

Domínio Processo Biológico

	<i>Nostoc</i> sp. CENA543	<i>Nostoc punctiforme</i> PCC73102	<i>Nostoc</i> sp. ATCC 53789	<i>Nostoc</i> sp. GBBB01	<i>Nostoc</i> sp. NIES-2111
Processo biológico (GO: 0008150)	1	3	1	2	5
Processo metabólico (GO: 0008152)	1	3		4	2
Processo celular (GO: 0009987)	1				
Processo metabólico da droga (GO: 0017144)	1				
Processo de desenvolvimento (GO: 0032502)	1				2
Processo metabólico celular (GO: 0044237)	1				
regulação biológica (GO: 0065007)	1				
Processo metabólico de carboidratos (GO: 0005975)		2			
Processo metabólico de compostos aromáticos celulares (GO: 0006725)		2		2	

Processo celular (GO: 0009987)	2			
Processo metabólico da macromolécula (GO: 0043170)	2		2	2
Processo metabólico celular (GO: 0044237)	2			2
Processo metabólico de polissacarídeo (GO: 0005976)	1			
Processo metabólico do composto contendo nucleobase (GO: 0006139)	1	1	2	
Processo metabólico do DNA (GO: 0006259)	1		1	
Processo de modificação de proteína celular (GO: 0006464)			3	
Processo metabólico de RNA (GO: 0016070)			2	
Processo metabólico de ácido orgânico (GO: 0006082)			1	
Processamento de RNA (GO: 0006396)			1	
Transporte mediado por vesículas (GO: 0016192)		1		
Transporte (GO: 0006810)		1		
Localização (GO: 0051179)		1		
Estabelecimento de localização (GO: 0051234)		1		
Localização de organela (GO: 0051640)		1		
Localização celular (GO: 0051641)		1		

Domínio Função molecular

	<i>Nostoc</i> sp. CENA543	<i>Nostoc punctiforme</i> PCC73102	<i>Nostoc</i> sp. ATCC 53789	<i>Nostoc</i> sp. GBBB01	<i>Nostoc</i> sp. NIES-2111
Atividade de transferase (GO: 0016740)	2		2	2	2
Atividade de hidrolase (GO: 0016787)			2		1
Atividade catalítica (GO: 0003824)			1		
Atividade de peptidase (GO: 0008233)			1		1

Domínio Componente celular

	<i>Nostoc</i> sp. CENA543	<i>Nostoc punctiforme</i> PCC73102	<i>Nostoc</i> sp. ATCC 53789	<i>Nostoc</i> sp. GBBB01	<i>Nostoc</i> sp. NIES-2111
Região extracelular (GO: 0005576)	1				
Membrana (GO: 0016020))					1
Parte da célula ((GO: 0044464)					1

Tabela S6: Metabólitos secundários identificados pelo antiSMASH no genoma da *Nostoc* sp. GBBB01.

Região	Tipo	Agrupamentos	Similaridade
Região 1.1	lanthipeptídeo, bacteriocina		
Região 3.1	hglE-KS, T1PKS	glicolipídeos heterocistos (KC489223.1.)*	85%
Região 4.1	Transat-PKS, T3PKS, PKS-like , T1PKS , NRPS	nosperina (GQ979609.2.)*	100%
Região 5.1	hglE-KS	glicolipídeos heterocistos (KC489223.1.)*	57%
Região 6.1	NRPS-like	nostoficina (JF430079.1.)*	36%
Região 6.2	NRPS	tiacumicina B (GQ979609.2.)*	6%
Região 8.1	bacteriocina		
Região 10.1	bacteriocina		
Região 11.1	Peptídeo lineares contendo azol (in)		
Região 12.1	terpeno		
Região 13.1	NRPS		
Região 20.1	lantidina		
Região 21.1	NRPS		

Região 22.1	terpeno		
Região 22.2	LAP, cianobactina	tricamida (NZ_AABK04000003.1)*	45%
Região 23.1	NRPS, T1PKS	nostopeptolídeo A2 (AF204805.2.)*	50%
Região 25.1	NRPS		
Região 28.1	terpeno		
Região 32.1	fosfonato		
Região 34.1	NRPS-like	anabaenopeptina NZ857 / nostamida A (CP001037.1.)*	100%
Região 34.2	microviridina	microviridina K (EU438895.1.)*	25%
Região 39.1	lantidina		
Região 50.1	lantidina		
Região 73.1	terpeno		

* Entrada originalmente do NCBI GenBank

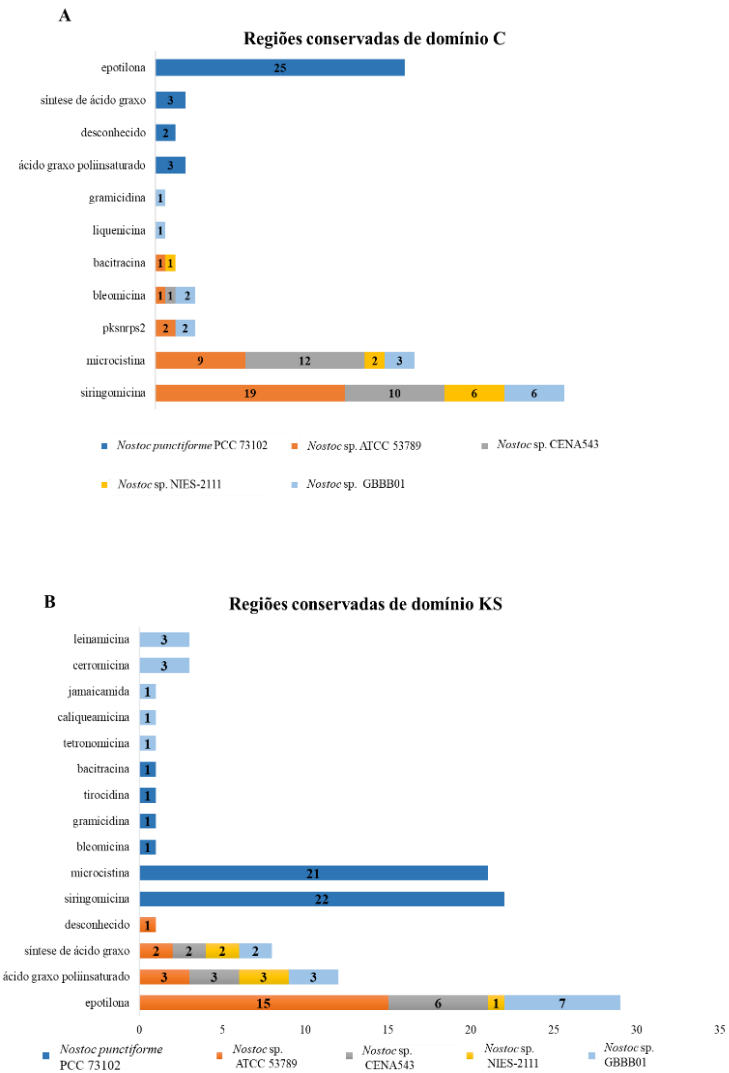


Fig. S1. (A) Classificação do NAPDOS referente às regiões de domínio C e KS identificado em cada genoma. (A) regiões de domínio C e (B) regiões de domínio KS

Tabela S7. Agrupamentos gênicos biossintéticos (BGC) identificados em *Nostoc* sp. GBBB01 pelo servidor NaPDos para domínio C

ID	Database Match ID	Identidade (%)	Alinhamento	e-value	Domínio	Classe	BLAST Proteína	Identidade %
C1	syri1_C9_LCL	43	297	1,00E-74	siringomicina	LCL	proteína contendo domínio de adenilação de aminoácido [<i>Nostoc punctiforme</i>]	99,66%
C2	syri1_C5_LCL	44	299	2,00E-68	siringomicina	LCL	proteína contendo domínio de adenilação de aminoácido [<i>Nostoc punctiforme</i>]	99,32%
C3	micro2_C1_DCL	44	299	2,00E-74	microcistina	DCL	peptídeo sintetase não ribossomal [<i>Nostoc punctiforme</i>]	99,66%
C4	grami2_C3_LCL	28	283	3,00E-22	gramicidina	LCL	proteína contendo domínio de adenilação de aminoácido [<i>Nostoc punctiforme</i> FACHB-252]	100%
C5	syri1_C6_LCL	44	297	1,00E-69	siringomicina	LCL	proteína contendo domínio de adenilação de aminoácido [<i>Nostoc punctiforme</i>]	100%
C6	micro2_C1_DCL	49	298	1,00E-87	microcistina	DCL	proteína contendo domínio de adenilação de aminoácido [<i>Nostoc punctiforme</i>]	100%
C7	micro1_C2E	47	300	6,00E-72	microcistina	epim	Oxidoreductase dependente de NAD (P) da família SDR [<i>Nostoc punctiforme</i>]	100%
C8	syri1_C9_LCL	41	296	1,00E-70	siringomicina	LCL	proteína contendo domínio de adenilação de aminoácido [<i>Nostoc punctiforme</i>]	97,96%
C9	syri1_C7_LCL	46	300	3,00E-73	siringomicina	LCL	proteína contendo domínio de adenilação de aminoácido [<i>Nostoc punctiforme</i>]	98,98%
C10	syri1_C5_LCL	46	300	3,00E-69	siringomicina	LCL	proteína contendo domínio de adenilação de aminoácido [<i>Nostoc punctiforme</i>]	100%

C11	bleom9_C2_cyc	41	295	3,00E-60	bleomicina	cíclico	proteína contendo domínio de adenilação de aminoácido [<i>Nostoc punctiforme</i>]	99,66%
C12	bleom9_C2_cyc	46	297	1,00E-69	bleomicina	cíclico	proteína contendo domínio de adenilação de aminoácido [<i>Nostoc punctiforme</i>]	99,64%
C13	Sare2407_1	37	296	1,00E-60	pksnrps2	C	proteína contendo domínio de adenilação de aminoácido [<i>Nostoc punctiforme</i>]	99,66%
C14	liche1_C3_LCL	34	295	2,00E-46	liquenicina	LCL	proteína contendo domínio de adenilação de aminoácido [<i>Nostoc punctiforme</i>]	98,29%
C15	Sare2407_1	29	295	6,00E-42	pksnrps2	C	proteína contendo domínio de adenilação de aminoácido [<i>Nostoc</i> sp. 2RC]	98,98%

Tabela S8. Agrupamentos gênicos biossintéticos (BGC) identificados em *Nostoc* sp. GBBB01 pelo servidor NaPDos para domínio KS

ID	Database Match ID	Identidade (%)	Alinhamento	e-value	Domínio	Classe	BLAST Proteína	Identidade %
KS1	PfaC_Shewanella_PUFA	46	457	3,00E-115	ácido graxo poliinsaturado	PUFA	Proteína da família PfaB [<i>Nostoc punctiforme</i>]	99,78%
KS2	PfaA_Shewanella_PUFA	65	458	0	ácido graxo poliinsaturado	PUFA	proteína contendo domínio de aciltransferase [<i>Nostoc punctiforme</i>]	99,78%
KS3	EpoC_Q9L8C8_H	59	427	4,00E-144	epotilona	HybridKS	proteína contendo domínio de aciltransferase [<i>Nostoc punctiforme</i>]	99,53%
KS4	EpoC_Q9L8C8_H	61	428	1,00E-162	epotilona	HybridKS	Oxidoreductase dependente de NAD (P) da família SDR [<i>Nostoc punctiforme</i>]	99,77%
KS5	EpoC_Q9L8C8_H	59	427	1,00E-144	epotilona	HybridKS	Oxidoreductase da família SDR [<i>Nostoc punctiforme</i>]	99,76%
KS6	EpoC_Q9L8C8_H	63	427	1,00E-164	epotilona	HybridKS	Oxidoreductase dependente de NAD (P) da família SDR [<i>Nostoc</i> sp. 2RC]	100%
KS7	FabF_Bacillus_FAS	34	446	1,00E-61	síntese de ácido graxo	FAS	beta-cetoacil-ACP sintase [<i>Nostoc punctiforme</i>]	99,04%
KS8	FabF_Bacillus_FAS	59	407	5,00E-138	síntese de ácido graxo	FAS	MULTISPECIES: beta-cetoacil-ACP sintase II [<i>Nostoc</i>]	99,51%
KS9	EpoC_Q9L8C8_H	62	427	9,00E-152	epotilona	HybridKS	Oxidoreductase dependente de NAD (P) da família SDR [<i>Nostoc punctiforme</i>]	100%
KS10	CALO5_12183629_i	27	431	2,00E-36	caliqueamicina	iterativo	Proteína da família PfaB [<i>Nostoc punctiforme</i>]	98,78%
KS11	PfaA_Shewanella_PUFA	65	456	6,00E-176	ácido graxo poliinsaturado	PUFA	Oxidoreductase dependente de NAD (P) da família SDR	100%

KS12	JamG_AAS98778_mod	49	224	1,00E-62	jamaicamida	modular	[<i>Nostoc punctiforme</i>] policetídeo beta-cetoacil: ACP sintase [<i>Nostoc</i> sp. 2RC]	100%
KS13	EpoC_Q9L8C8_H	60	427	1,00E-145	epotilona	HybridKS	Oxidoreductase dependente de NAD (P) da família SDR [<i>Nostoc punctiforme</i>]	99,30
KS14	KirAII_CAN89632_3T	48	435	8,00E-101	cerromicina	trans	proteína contendo domínio de adenilação de aminoácido [<i>Nostoc punctiforme</i>]	99,54%
KS15	KirAII_CAN89632_5T	46	437	7,00E-109	cerromicina	trans	proteína contendo domínio de adenilação de aminoácido [<i>Nostoc punctiforme</i>]	100%
KS16	LnmI_AF484556_2T	45	435	7,00E-102	leinamicina	trans	proteína contendo domínio de adenilação de aminoácido [<i>Nostoc punctiforme</i>]	100%
KS17	EpoC_Q9L8C8_H	43	425	1,00E-85	epotilona	HybridKS	proteína contendo domínio de adenilação de aminoácido [<i>Nostoc punctiforme</i>]	99,76%
KS18	LnmJ_AF484556_4T	36	436	2,00E-78	leinamicina	trans	proteína contendo domínio de adenilação de aminoácido [<i>Nostoc punctiforme</i>]	99,53
KS19	LnmJ_AF484556_2T	51	427	9,00E-121	leinamicina	trans	Oxidoreductase dependente de NAD (P) da família SDR [<i>Nostoc punctiforme</i>]	99,76
KS20	KirAIV_CAN89634_7T	52	439	3,00E-126	cerromicina	trans	Oxidoreductase dependente de NAD (P) da família SDR [<i>Nostoc punctiforme</i>]	99,77%
KS21	TetH_BAE93740_1T	49	436	1,00E-114	tetronomicina	trans	Oxidoreductase dependente de NAD (P) da família SDR [<i>Nostoc punctiforme</i>]	100%

