

TITLE: CHARACTERIZATION AND PROSPECTING OF THE GENOME OF THE CYANOBACTERIUM *NOSTOC* SP. GBBB01 FROM CERRADO BIOME USING A COMPUTATIONAL APPROACH

AUTHORS: COSTA, A.L.S.¹; ROCHA, I.K.B.S.¹; BARROS, J.I.P.¹; BUTARELLI, A.C.A.²; FERREIRA, L.S.S.²; LIMA, A.R.J.³; DALL'AGNOL, H.¹; DA SILVA, A.M.²; SETUBAL, J.C.²; DALL'AGNOL, L.T.¹

INSTITUTIONS: ¹UNIVERSIDADE FEDERAL DO MARANHÃO, SÃO LUÍS, MA (AVENIDA DOS PORTUGUESES, 1966, CEP 65080-805, SÃO LUÍS – MA, BRAZIL). ²UNIVERSIDADE DE SÃO PAULO, SÃO PAULO, SP (Av. Prof. Lineu Prestes, 748 - Butantã, São Paulo - SP, 05508-900, SÃO PAULO – SP, BRAZIL). ³INSTITUTO BUTANTAN, São Paulo – SP, (Av. Vital Brasil, 1500 - Butantã, São Paulo - SP, 05503-900, São Paulo – SP, Brazil).

ABSTRACT:

Cyanobacteria have phenotypic and ecological plasticity, which allows them to adapt to different environments and regions. The cyanobacteria community of the Chapada das Mesas National Park - Maranhão (PNCM) is ecologically important and has biotechnology potential but is poorly characterized. Therefore, the objective of this work was to improve the genome annotation, especially of hypothetical proteins (HPs), and perform the genome mining of the *Nostoc* sp. GBBB01 non-axenic lineage originally isolated from the PNCM and deposited at the GBBB collection at UFMA and sequenced at the Center for Advanced Technologies in Genomics at Chemistry Institute from USP. After metagenomic sequencing, de novo assembly was performed with the Metawrap Assembly module using metaSPAdes v. 3.13. Assembled contigs were subjected to three different binning rounds using CONCOCT, MaxBin2, and MetaBAT2. The annotation of the isolated genome of the GBBB01 strain was performed using the Prokaryotic Genome Annotation Pipeline (PGAP) and RAST. In the structural and comparative characterization of the genome, antiSMASH 6.0, NapDoS, PHASTER, CRISPRCasFinder, TYGS, GGDC and MEGAX tools were used. Manual curation of the annotation of gene clusters was performed using the computational tool BLAST v2.6.0+ (NCBI), PFAM, UniProtKB, HAMAP and its visualization by Artemis. The analysis of hypothetical proteins targeted their physicochemical properties and sub-cellular localization, using sequence comparison by BLAST in ProtParam, PSORTb, GRAVY, and STRING databases. The Neighbor Joining analysis of 16S rRNA confirmed the relationship of the *Nostoc* sp. GBBB01 with other strains of the *Desmonostoc* and *Nostoc* genera. As for the functional analysis, 25 functional categories were identified by the RAST server and 24/25/21 gene groupings were identified by AntiSMASH 6.0 and NapDoS (C and KS), respectively related to terpenes, microviridines, bacteriocins, microcystins and other biosynthetic pathways of the NRPS/PKS/RiPP's types. A total of 114 HPs related to biosynthetic gene clusters were identified and had their annotation improved through several online tools. The analysis revealed genes related to toxins with relevance to one health in an important touristic site in the state, in addition to the great potential in the production of bioactive compounds relevant to the proper valorization of the biodiversity of a threatened biome.

Keywords: Bioinformatics; Cyanobacteria, Hypothetical proteins.

Development agency: Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq), Fundação de Amparo à Pesquisa e ao Desenvolvimento Científico e Tecnológico do Maranhão (FAPEMA) and Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES).