

**TITLE:** GENOME MINING AND ANNOTATION OF THE CYANOBACTERIA *SYNECHOCYSTIS* CACIAM 05

**AUTHORS:** RIBEIRO, I.S.<sup>1</sup>; LIMA, A.R.J.<sup>2</sup>; GONÇALVES, E.C.<sup>2</sup>; MORAES, P.H.G.<sup>1</sup>; DALLAGNOL, H.P.M.B.<sup>1</sup>; DALLAGNOL, L.T.<sup>1</sup>

**INSTITUTION:** <sup>1</sup>UNIVERSIDADE FEDERAL DO MARANHÃO, SÃO LUÍS, MA (AV. DOS PORTUGUESES, 1966 - VILA BACANGA, SÃO LUÍS - MA, 65065-545). <sup>2</sup>UNIVERSIDADE FEDERAL DO PARÁ, BELÉM, PA (RUA AUGUSTO CORRÊA, 01 - GUAMÁ, BELÉM - PA, 66075-110)

#### **ABSTRACT**

As cyanobacteria have a vast distribution throughout the planet where the most varied ecosystems are colonized and in which they are large ecological and economic dimensions. In addition, this group has great relevance in the area of biotechnology because it has potential in the production and variety of secondary metabolites and bioactive compounds. The use of computational tools has been promising in the formation of databases, with the development of algorithms that perform the search and automatic evaluation of biosynthetic pathways grouping of secondary metabolic. The objective of this work is to perform an manual annotation and mining of the genome of the *Synechocystis* sp. CACIAM 05. The genome was sequenced by the Illumina MiSeq platform by the Biomolecular Technology Laboratory/UFGA. The automatic annotation was performed using MG-RAST server and the NCBI Prokaryotic Genome Annotation Pipeline. A manual curation of the automatic annotation was carried out using the BLAST v2.6.0 + (NCBI), PFAM, UniProtKB, HAMAP computational tool and visualized in Artemis. The automatic annotation identified 3,327 coding regions with 1,142 classified as hypothetical proteins. For gene clusters and biosynthetic pathways analysis the genome was submitted to the servers NapDos and NPsearch and for the antiSMASH version 4.0 with the genome in FASTA, without external annotations, and the following options: with the options enabled: (i) KnownClusterBlast; (ii) smCoGanalysis; (iii) Align Trans-AT PKS domains; (iv) ClusterBlast; (v) ActiveSiteFinder; (vi) Whole-genome PFAM analysis; (vii) SubClusterBlast; (viii) Detect TTA coódon. AntiSMASH detected 22 probable gene clusters in which 5 clusters presented hypothetical protein genes as a main biosynthetic gene. Three clusters are of terpene type that presented 52%, 48% and 54% of gene similarity with the groupings *Synechocystis* sp. PCC 6803. Two clusters of fatty acids were identified. Clusters related to fatty acid biosynthesis presented 28% similarity with *Leptolyngbya* sp. PCC 7375 and 7% with *Synechococcus* WH7803, respectively. The NP researcher found two genes of the mevalonate-independent biosynthetic pathway that is used in the production of isoprenoids in eukaryotes and many bacteria. These preliminary results are an attempt to optimize the genomic characterization of CACIAM 05, as it will be used as a base for secondary metabolites prospection and biotechnological exploitation of this organism.

**Keywords:** Automated and manual annotation, *in silico*, hypothetical proteins

**Development Agency:** CAPES; CNPq; FAPESPA; FAPEMA.